Project Thesis

Development and Evaluation of an Algorithm for Spatial Enhancement of Ambisonic Room Impulse Responses

Elias Hoffbauer

Supervisor: Dr. Matthias Frank

Graz, February 9, 2021

institut für elektronische musik und akustik

Elias Hoffbauer (Name in Blockbuchstaben)

01473027 (Matrikelnummer)

Erklärung

Hiermit bestätige ich, dass mir der *Leitfaden für schriftliche Arbeiten an der KUG* bekannt ist und ich die darin enthaltenen Bestimmungen eingehalten habe. Ich erkläre ehrenwörtlich, dass ich die vorliegende Arbeit selbständig und ohne fremde Hilfe verfasst habe, andere als die angegebenen Quellen nicht verwendet habe und die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Graz, den 09.02.2021

E. Hoffbauer

Unterschrift der Verfasserin/des Verfassers

Zusammenfassung

Für die immersive Nachbildung eines Raumklangs mit z. B. einem Faltungshall bietet es sich an, eine Ambisonische Raum-Impulsantwort (Ambisonic room impulse response, ARIR) mit einem Mikrofonarray zu messen. Diese wird üblicherweise aus finanziellen und praktischen Gründen in der ersten Ambisonischen Ordnung (FOA) aufgenommen. Höhere Ambisonische Ordnungen (HOA) versprechen wiedergabeseitig dagegen eine genauere Richtungsauflösung und bessere Tiefenstaffelung und somit eine größere, subjektiv empfundene Natürlichkeit des Klangerlebnisses.
Daher ist es ein vielverspechender Ansatz, Algorithmen zu entwerfen, die ausgehend von aufgenommenen Signalen niedriger Ordnungen ein möglichst naturgetreues Ambisonisches Signal für höhere Ordnungen rekonstruieren und so die Vorteile der Aufnahme- und Wiedergabe-Domäne miteinander verbinden.
Im ersten Teil dieser Projektarbeit wird ausgehend von der bestehenden Spatial Decomposition Method ein Algorithmus entwickelt, der aufbauend auf der Schätzung des Pseudo-Intensitätsvektor mehrere Richtungen des Ausgangssignals dekodiert und daraus eine Impulsantwort beliebig hoher Ordnung rekonstruiert. Im zweiten Teil werden anschließend seine klanglichen Vorteile gegenüber anderen Algorithmen mit einem Hörversuch untersucht und ausgewertet.

Abstract

For the immersive auditory representation of a room, e. g. with a convolution reverb, it is useful to measure an Ambisonic room impulse response (ARIR) with a microphone array. This is usually performed in first-order Ambisonics (FOA), out of practical and monetary reasons.
However, from the playback perspective higher-order Ambisonics (HOA) has many advantages, like a sharply resolved representation of directions and depth, which results in a recording, that is perceived as very natural sounding. It is a promising approach to develop algorithms that enhance measured signals to a higher order in the most realistic way as possible and combine in that way the advantages of both the recording and playback domain.
In the first part of this project thesis an algorithm is developed based on the principles of the Spatial Decomposition Method. It decodes multiple directions of a given first-order signal via the estimation of the pseudo-intensity vector and encodes them again in any desired order. In the second part the algorithm is compared to other known algorithms in a listening test and advantages are investigated.

Contents

1 Rendering Impulse Responses in HOA: Motivation and Problems

In many post-production scenarios, it is useful being able to change the spatial impression of a recording, or in general of a signal, for example adding some reverberation to an orchestra recording that sounds too dry or creating a soundscape for a virtual reality application. One way to do that, is using measured impulse responses and convolve them with the recorded or synthesized signal. The signal is enriched with the reverberation of the room, in which the impulse response measuring was performed. One advantage of this method is the realistic, very natural, and convincingly sounding result, without having to adjust a lot of parameters for synthesis of reverberation with a feedback delay network or similar [SP82].
In spatial audio, especially in the Ambisonics domain, this technique is very appealing, because now as well the direction of the direct sound and early reflections of the original room can be reproduced in the playback of the convolved signal over a multichannel loudspeaker setup or binaurally over headphones.
Ambisonics does not simply enhance the immersive experience, but also demands higher requirements: The recorded impulse response has to be recorded with a microphone array, that allows converting the recorded signals of the multiple microphone capsules into Ambisonic signals $\boldsymbol{\chi}_N$ described in the spherical harmonics domain. The order N of the Ambisonic signals is determined by the number of the capsules; with C capsules the maximum order is $N = \sqrt{C} - 1$.
On the recording side, using microphone arrays of low order is often preferred. They are rather low-cost and allow a higher sound quality. The conincidence constraint limits the size of the array, but can be fulfilled more easily in a low order with an arguable adequate membrane radius. Further, fewer capsules generate lower noise floor.
On the other hand, at the side of playback, a higher order yields several advantages: The sweetspot gets bigger, up to 2/3 of the loudspeaker radius for fifth order and the perceived spatial depth increases [FZ16, FZ17]. Also, the resolution of directions gets much more sharper and the source width decreases.
With the development of an algorithm for this *upmixing* application, it is possible to combine the advantages of FOA-recording and HOA-reproduction. The preservation of the coloration and spatial properties of the original ARIR is crucial for a well functioning method. Tervo et al. could show in listening tests, that SDM changes the timbre of impulse response less than the *Spatial Impulse Response Rendering* (SIRR), developed by Pulkki et al. [TPKL13, PML04]. However, a major drawback of SDM compared to SIRR are the prominent artefacts, which arise during the upmixing process. They are clearly

audible in the unconvolved ARIR and can also appear after convolution with a very transient signal.
It is the motivation of this project thesis to develop a new upmixing technique based on SDM, which works as a tool for enhancing FO-ARIRs to an arbitrary chosen Ambisonic order, preserves all properties of the original impulse response and contains as less artefacts as possible.

2 First Approach: SDM with Ambisonic Widening

The *Spatial Decomposition Method* estimates the DOA of the sound for every sample based on the measured signals of the capsules in microphone array (s. sec.3.1). In a second step, the omnidirectional signal is encoded samplewise in the estimated direction. Applying *Ambisonic Widening* [ZF19, p.121f; ZFKC14] on the SDM-upmixed impulse response was the first approach: The sound components of the Ambisonic signal are panned away from their original position on defined trajectories, the degree is depending on their frequency (s. Fig. 1). This leads to a perceived source widening of the individual sound events, which could in theory mask the audible artefacts and let them appear less prominent.

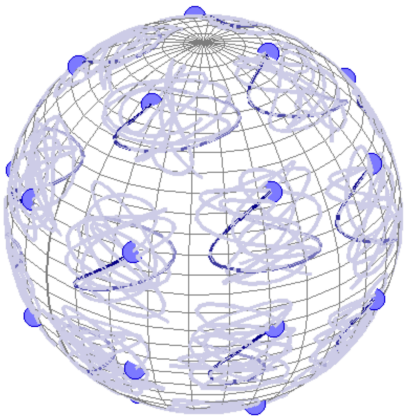

Figure 1 – Trajectories for frequency-dependent panning on a sphere, from [ZFKC14].

It seems necessary to separate the direct from the diffuse part of the ARIR and treat them individually, because Ambisonic Widening of the direct sound and the early reflections evokes both spatial and temporal smearing of the measured impulses, which is the opposite of what is desired to achieve using the upmixing procedure.
The separation is realised with the peak-picking method proposed by Müller [Mül19]; different widths and minimal distances of peaks were tested, as well as several thresholds for a minimum height of single peaks. Detected peaks are stored in the direct-signal vector, the rest in the vector for the diffuse part. Further different forms of slopes and the fading lengths between direct and diffuse part at merging after the widening were examined.
Informal listening tests of the author show that even with carefully chosen parameters the artefacts are still clearly perceivable and a change in coloration of the ARIR arises due to the comb filters, which are a side effect of the widening. Also, the choice of parameters depends strongly on the properties of the ARIR, e.g. if it is very reverberant or rather dry, so it is hardly possible to find a set of parameters that can be applied sufficiently on a wide range of very different room impulse responses.
For determining what the best achievable result of artefact masking with this method would be, the whole impulse response is completely widened. The still existing smearing of the impulse in the direct part was ignored for a start. This variant was also presented in both parts of the listening experiment (s. sec. 4).

3 The 4DE-Algorithm

The fundamental idea of the 4DE-algorithm is to gather more signal information from recorded Ambisonic room impulse responses, than the existing classic SDM does, which encodes only the omnidirectional channel $\boldsymbol{h}_0^0$ in one estimated direction [TPKL13]. Using the 4DE-algorithm, four directions in a tetrahedral arrangement are decoded samplewise via four beamformers and again encoded in that same direction in a higher Ambisonic order that can be chosen freely. In the documentation, as well in the code of the algorithm, the ambiX-convention is obeyed, that is the *Ambisonic Channel Numbering* (ACN) and SN3D-weighting [NZDS11]. A list of variables used in this section can be found in the Appendix A.

3.1 Estimation of the Direction of Arrival (DOA)

The DOA estimation is performed similar to SDM [TPKL13]: In order to prepare the FO-ARIR $\boldsymbol{h}_N$ with $N = 1$, it is bandpass filtered from 200 Hz to 4 kHz, because very low and very high frequencies introduce effects of diffraction and interference and possibly

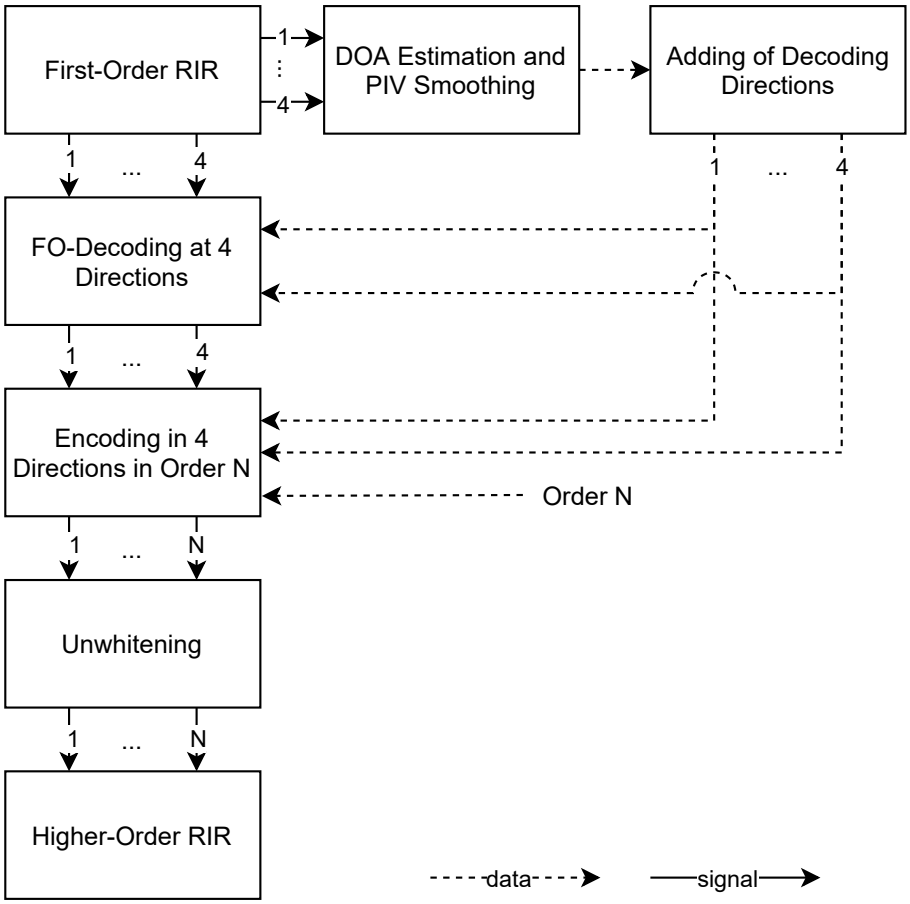

Figure 2 – Flow chart of the developed 4DE-algorithm.

worsen the results causing spatial aliasing. For low frequencies, very little pressure differences get registered between the capsules of the array, whereas high frequencies get reflected at the rigid array surface.
For each sample k the DOA of the sound is estimated. A very efficient way of implementation is the Pseudo Intensity Vector (PIV), named $\boldsymbol{I}$ in here. It is samplewise computed through the multiplication of the omnidirectional first channel $\boldsymbol{h}_0^0$ of the ARIR, also called the W-channel after the FuMa-Convention [Mal99], and the three remaining channels, $\boldsymbol{h}_1^{-1}$ (Y-), $\boldsymbol{h}_1^0$ (Z-) and $\boldsymbol{h}_1^1$ (X-channel) of the first Ambisonic order. Its orientation is equivalent to the negative DOA and its length proportional to the intensity from that direction

$$\boldsymbol{I}[k] = W[k] \begin{bmatrix} X[k] \\ Y[k] \\ Z[k] \end{bmatrix} = \boldsymbol{h}_0^0[k] \begin{bmatrix} \boldsymbol{h}_1^1[k] \\ \boldsymbol{h}_1^{-1}[k] \\ \boldsymbol{h}_1^0[k] \end{bmatrix} \propto \boldsymbol{h}_0^0[k]\boldsymbol{v}[k]\,. \tag{1}$$

Although two pressure values are multiplied, it is a good approximation for the multiplication of $\boldsymbol{h}_0^0$ and its particle velocity vector $\boldsymbol{v}$.
In the implementation, the resulting vector $\boldsymbol{I}$ was smoothed with a median window of 10 samples length, which equals approximately 0.2 ms at a sampling frequency of 48 kHz. This value was chosen for an optimal signal detection, regarding the largest distance between two capsules of the first-order microphone, as recommend by Lokki in [TPKL13].

3.2 Adding of Decoding Directions

Based on the direction as estimated sec. 3.1, three more directions are added. In the first step, direction 2 is calculated by tilting the estimated direction 1 downwards by 109.5°. Direction 3 and 4 are calculated by rotating direction 2 by 109.5° clockwise and anticlockwise. The rotation axis is defined by the orientation of the unity-vector of direction 1. With these subsequent rotations, one gets four directions in a regular tetrahedral arrangement. Fig. 3 illustrates the process of rotations graphically.

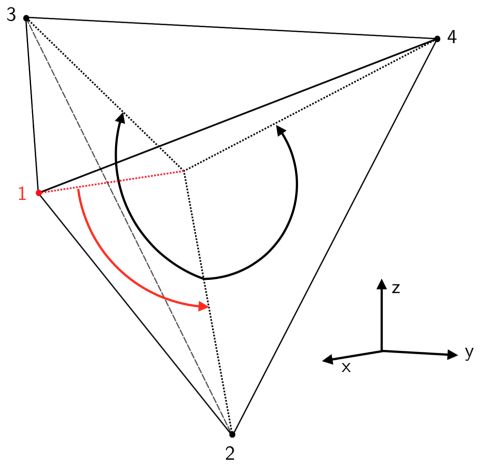

Figure 3 – Added directions (black) based on the estimated direction (red).

3.3 Decoding of Signals

Geometry Decoding a direction of an Ambisonic scene with a Sampling-Decoder (SAD) equals recording sound in the real soundfield with an ideal hyper-cardioid microphone aligned in that direction. The four directions are distributed equally on the sphere and one is able to get four completely decorrelated signals, because every decoding position lies in the zeros of the hyper-cardioid directivity pattern of its three neighbouring positions. The directivity D of a hyper-cardioid as a function of the direction θ on a horizontal plane in two dimensions is described as following:

$$D(\theta) = 0.25 + 0.75\cos(\theta)\,. \tag{2}$$

Calculating its zeros with $D(\theta) = 0.25 + 0.75\cos(\theta) \stackrel{!}{=} 0$, we get

$$\theta = \arccos\left(-\frac{1}{3}\right) \approx 109.5° \tag{3}$$

which equals the angle α between lines from the center to any two vertices of a regular shaped tetrahedron [WTFE20].
In the following two figures 4 and 5, the link between the directivity zeros and the tetrahedral directions is shown graphically. The grey dotted lines lie in the $0°$ direction, the red dotted lines point in the directions of the zeros of the hyper-cardioid directivity pattern. Because it is a regular tetrahedron, it can be demonstrated, that this relation is valid for every single direction.

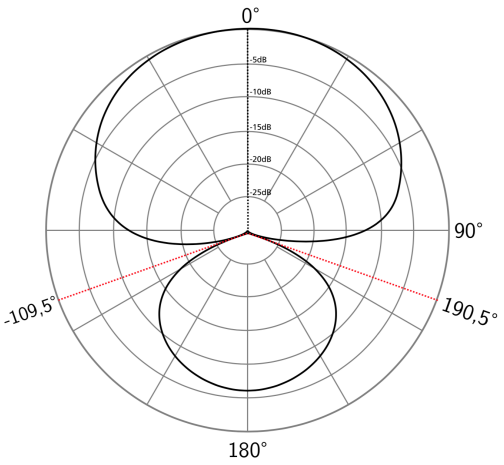

Figure 4 – 2-dim. directivity pattern of a hyper-cardioid microphone.

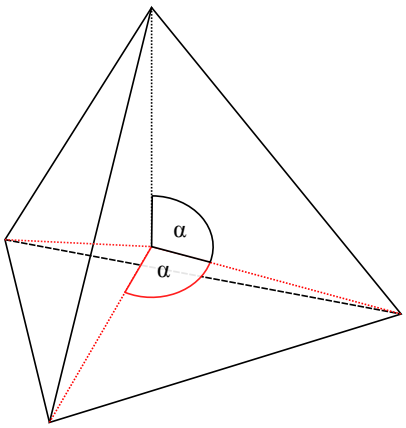

Figure 5 – Two exemplary angles in a regular tetrahedron.

Signal Processing The four directions are decoded from the first-order Ambisonics room impulse response $\boldsymbol{h}_{N=1}$ by multiplication with a samplewise computed decoder matrix $\boldsymbol{D}$ of the SAD with the dimensions 4×4. It consists of the matrix $\boldsymbol{Y}_N^T$, which contains the weights of the spherical harmonics evaluated in the four directions θ_1 to θ_4 and the correction factor π, which yields from decoding not one, but four directions. The four signals are stored in $\boldsymbol{S} = \begin{bmatrix}\boldsymbol{S}_1 & \boldsymbol{S}_2 & \boldsymbol{S}_3 & \boldsymbol{S}_4\end{bmatrix}^T$ after the multiplication of the decoder with the Ambisonic input signals $\boldsymbol{\chi}_{N,in}$.

$$\boldsymbol{S} = \boldsymbol{D}\boldsymbol{\chi}_N = \pi\, \boldsymbol{Y}_1^T \boldsymbol{\chi}_{1,in} = \begin{bmatrix}\boldsymbol{y}_1(\theta_1) & \dots & \boldsymbol{y}_1(\theta_4)\end{bmatrix}^T \boldsymbol{\chi}_{N,in}\,. \tag{4}$$

Equation 4 depicts the calculation executed for each set of directions at one sample in time.

3.4 Encoding of Signals

Signal Processing The encoding of the signals is executed similarly to the decoding. By rearranging the decoding equation 4, one gets the encoding equation:

$$\boldsymbol{\chi}_{N,out} = \boldsymbol{D}^T \boldsymbol{S} = \boldsymbol{Y}_N \boldsymbol{S} = \begin{bmatrix}\boldsymbol{y}_N(\theta_1) & \dots & \boldsymbol{y}_N(\theta_4)\end{bmatrix} \boldsymbol{S}\,. \tag{5}$$

The essential step of upmixing happens during the encoding of the signals. The four signals are encoded again in the same direction, they were decoded in the computational step before, but now with a higher spatial resolution, depending on the chosen Ambisonic order N. Now, $(N+1)^2$ spherical harmonics are evaluated at $\theta_1 \dots \theta_4$ and stored in the weighting matrix $\boldsymbol{Y}_N$. $\boldsymbol{D}$ has now the dimensions $(N+1)^2 \times 4$. The Ambisonic signal $\boldsymbol{\chi}_{N,out}$ can now be used as upmixed impulse response $\tilde{\boldsymbol{h}}_N$.

3.5 Spectral Correction: Unwhitening

Through the fast angular changes of the pseudo-intensity vector, a strong amplitude modulation is introduced at higher Ambisonic orders during the encoding process. Especially in the late reverberation tail of the impulse response, where the sound field reaches the diffuse state and the sound comes equally distributed from every direction. The signal becomes more transient and the level of high frequencies increases. This effect is also clearly audible as coloration of the upmixed ARIR, which gets brighter than the original. The equalisation of this spectral distortion, called *Unwhitening*, is conducted in accordance with Zotters and Franks method, described in detail at [ZF19, p.125f]: A reference RMS-value for every third-octave band b at a discrete point in time k is computed, and the upmixed Ambisonic signals in $\tilde{\boldsymbol{h}}_n^m$ are corrected by a factor, which is constant through the order n. A small adaptation was made, the basis for the reference calculation is not the only the W-channel $\boldsymbol{h}_0^0$ any more, but a mixture to same parts of the W-channel and the signals of the first order:

$$\boldsymbol{h}_n^m[k,b] = \tilde{\boldsymbol{h}}_n^m[k,b] \sqrt{\frac{(2n+1)\mathcal{E}\{|\frac{1}{2}\boldsymbol{h}_0^0[k,b] + \frac{1}{6}\sum_{m=-1}^{1} \boldsymbol{h}_1^m[k,b]|^2\}}{\sum_{m=-n}^{n} \mathcal{E}\{|\frac{1}{2}\tilde{\boldsymbol{h}}_m^n[k,b]|^2\}}}\,. \tag{6}$$

With this equation, one gets the equalised impulse response $\boldsymbol{h}_N[k, b]$. $\mathcal{E}\{|\cdot|^2\}$ estimates the squared signal envelope.

3.6 Possible Preservation of the First-Order Ambisonic Signal

In theory, it is possible to preserve the first-order signal throughout the en- and decoding process. A regular tetrahedron is a optimal spherical t-design with L = 4 points and degree $\tau(L)$ = 2 [SH02]. For Ambisonic room impulse response $\boldsymbol{h}_N$ of order $N = 1$, which satisfies $t(L) > 2N$, orthonormality is guaranteed and it can be deduced that

$$\boldsymbol{D}\boldsymbol{D}^T = \frac{S_D}{L}\boldsymbol{Y}_N\boldsymbol{Y}_N^T = \boldsymbol{I}_N \tag{7}$$

[ZF19, p.80], with the surface of the unit sphere in three dimensions S_D and the number of directions L.
To find the link between the input and output first-order signals, we use equations 4 and 5 from section 3.3 and 3.4. After inserting one into the other, we get

$$\boldsymbol{\chi}_{\tilde{N},out} = \pi\,\boldsymbol{Y}_1\boldsymbol{Y}_1^T\boldsymbol{\chi}_{1,in}\,. \tag{8}$$

Only the signals up to the first order of the output Ambisonic signals are of interest. So in order to ensure dimensional consistency, the weighting matrix $\boldsymbol{Y}_N$ of the encoding step was trimmed back to the dimension 4×4 and thus equals $\boldsymbol{Y}_1$. The back-trimmed Ambisonic signal $\boldsymbol{\chi}_{\tilde{N},out}$ is defined as subvector of the full output vector $\boldsymbol{\chi}_{N,out}$, which contains the k-th sample all channels of the impulse response $\boldsymbol{h}_1$ that was upmixed to order N.

$$\boldsymbol{\chi}_{N,out}[k] = \big[\big[\boldsymbol{\chi}_{\tilde{N},out}[k]\big] \quad \dots \quad \boldsymbol{h}_N^N[k]\big] = \big[\big[\boldsymbol{h}_0^0[k] \quad \dots \quad \boldsymbol{h}_1^1[k]\big] \quad \dots \quad \boldsymbol{h}_N^N[k]\big]\,. \tag{9}$$

$\boldsymbol{h}_n^m$ are column vectors of the measured Ambisonic room impulse response of order n, degree m and length of K. In the case of constant encoding and decoding directions, K equals the length of the room impulse response. In the application of the 4DE-algorithm, the estimated and added directions are constantly changing from sample to sample and thereby also the weighting matrix $\boldsymbol{Y}$ changes. Consequently $K = 1$, because each sample is en- and decoded with its individual decoding-matrix $\boldsymbol{D}$. Equation 7 with the parameters $S_D = 4\pi$ and $L = 4$, as applied in the 4DE-algorithm, simplifies equation 8 to

$$\boldsymbol{\chi}_{\tilde{N},out} = \boldsymbol{I}_1\boldsymbol{\chi}_{1,in}\,. \tag{10}$$

Equation 10 shows that the first order of the input and of the output signal are identical given the ideal constraints are fulfilled.
However, a numerical analysis of the algorithm indicates that this ideal results cannot be reached. A relative energetic error in dB is computed by Eq. 11 for each sample k:

$$\boldsymbol{E}[k] = 10\log\left(\frac{1}{4}\sum\frac{\Delta\boldsymbol{h}_0^0[k]}{\boldsymbol{h}_{0,in}^{0\ 2}[k]} + \dots + \frac{\Delta\boldsymbol{h}_1^1[k]}{\boldsymbol{h}_{1,in}^{1\ 2}[k]}\right) \tag{11}$$

with

$$\Delta \boldsymbol{h}_m^n[k] = |\boldsymbol{h}_{n,out}^{m\ 2}[k] - \boldsymbol{h}_{n,in}^{m\ 2}[k]| \,. \tag{12}$$

To avoid invalid results due to arithmetic operations with 0, the values of the rarely occurring samples where the denominator $\boldsymbol{h}_{n,in}^{m\ 2}$ reaches 0, are replaced with the first non-zero value from the sample $k - i$, i steps earlier in time. Samples, where $\Delta \boldsymbol{h}_n^m$ equals 0, are replaced by the value $1\mathrm{e}^{-6}$ to prevent the logarithm approaching $-\infty$.
The mean energetic error over time and all four signals, calculated with the ARIR measured in the St. Andrew's Church in Lyddington for 100 ms ($\approx$ 4800 samples at 48 kHz) beginning at the first impulse, is approximatly -27 dB. This equals also roughly the result for the energetic error of the impulse response upmix of the Minster in York. The biggest impact on the error has the rounding of the tetrahedral angle to 109.5°. Calculations using the exact angular value $\alpha = \arccos(-\frac{1}{3})$ show that the error can be minimised down to roughly -60 dB. Besides that, the finite dynamic resolution, in this case 24 bit, of the impulse response measurement and the numerical errors of MATLAB [1] have to be taken into account.
It is important to note, that these comparisons between input and output were made with upmixes, which were not equalised with Unwhitening. This procedure heavily magnifies the error even further up to approximatly -3 dB through manipulations in the individual frequency bands (s. sec. 3.5).
Additional steps for solutions of this problem were discussed during the project thesis, but would exceed the time and effort of this work. One idea would be to exclude the first-order of the upmixed Ambisonic signal from Unwhitening, since it is besides a negligible error the same as the input signal, and therefore has not to be spectrally corrected.

1. https://www.mathworks.com/products/matlab.html .

4 Listening Test

After the development of the 4DE-algorithm a listening experiment is conducted. The new algorithm is compared to other upmixing methods such as SDM with Widening (s. 2), the classic SDM [TPKL13], the *Higher-order Ambisonic Spatial Decomposition Method* (HO-ASDM) [Dep19] and the *Spatial Impulse Response Rendering* (SIRR) [PML04].

4.1 Test Design

The aim of the listening experiment is to investigate (I) how well the artefacts get cancelled out and (II) how accurately the coloration of the original ARIR is preserved. The parts are presented in a predefined order, the trials and stimuli are randomized.
In part I, impulse responses of three rooms with different reverberation times are presented. In trial 1, a ARIR measured in the St. Andrew's church in Lyddington[2] is used; in trial 2 a ARIR of the large Minster in York[2] and in the third trial a ARIR of the CUBE, a room for lectures of medium size at the iem in Graz. The measurement in the CUBE is performed with a Zylia M-1 [3] and afterwards rendered to the AmbiX-format with the *array2sh* plug-in from the *SPARTA* suite [4].
The subjects are asked to compare the strength of the artefacts of the different, unconvolved upmixing methods to the measured first-order Ambisonic room impulse response (FO-ARIR) as reference. In between the different stimuli the first-order measurement is again hidden, for the possibility to cancel out the invalid sets from persons, that did not even recognise the same sample. The pseudo-continuous scale spans from 0, *very dominant artefacts*, to 10, *no audible artefacts* and has a step size of 0.01.
In the second part, the same ARIRs are used, except in the last trial, where the impulse response of the CUBE is exchanged with a fourth-order impulse response measurement by an Eigenmike em32 [5] in the St. Paul's Concert Hall in Huddersfield [6]. The results of this trial should answer, how an impulse response measured for higher-order Ambisonics is assessed regarding coloration, compared to upmixed impulse responses and the first-order reference. The measured ARIRs are convolved with white noise for a broad-band excitation of all frequencies and the most transparent presentation of differences in the coloration. Only the steady-state part of the convolved audio files is presented to ensure a stable, invariant timbre. To obtain ratings, the subject are asked to rank the timbre of the stimuli, compared to the reference from *very different* to *identical* on a pseudo-continuous (interval size 0.01) scale from 0 to 10. It is well-known, that timbre is a multi-dimensional value. However, the interpretation and use of the different properties for the ranking are left to the subjects, because a subjective overall evaluation of the coloration differences is wanted in this part of trials.

2. provided by www.openairlib.net, Audiolab, University of York, Damian T. Murphy.
3. https://www.zylia.co/zylia-zm-1-microphone.html .
4. http://research.spa.aalto.fi/projects/sparta_vsts/plugins.html .
5. https://mhacoustics.com/products .
6. taken from the 3D-MARCo projekt, https://zenodo.org/record/3477602 .

4.2 Statistical Methods

First, the normal distribution for every stimulus is tested with the *Kolmogorov-Smirnov test*. Results show, that in none of the six trials all stimuli were normal distributed, so for the further investigations a non-parametric tests are used.
The *Kruskal-Wallis test* shows clearly, that there is in every trial at least one group, that differs significantly from the population. A pairwise comparison of every stimuli group is performed with the *Wilcoxon signed-rank test* for every trial. With the *Bonferroni-Holm method* probability values are adapted and the multiple comparisons problem is minimized. The full results for every trial can be found in Appendix B.

4.3 Results

16 well-trained male subjects, all master students or alumnis of the iem, took part in the experiment. Four persons did it twice, which led in a test strength of 20 sets of results. The average age was approximately 30 years.
The following plots show the median and the confidence interval (CI) with the confidence coefficient of 0.95.

4.3.1 Part I: Investigation of the Strength of Artefacts

The hidden reference FO-ARIR is recognised in every trial of part I, as seen at the significant best ratings ($p < 0.01$, s. Appendix B). The classic SDM is always valued significantly ($p < 0.01$) with lowest score. The first approach to mask the artefacts (SDM+Widening) proves to be as well a very significant improvement ($p < 0.01$) to the SDM regarding the artefacts, but is still assessed by trend worse than the other upmixing variants. A more detailed analysis follows in the individual paragraphs.

Trial 1 The ARIR of the church in Lyddington comprises a good amount of reverberation and has a broadband reverberation time T_{30} of approximately $1.4\,\mathrm{s}$. The 4DE-algorithm is significantly better rated than the other upmixing variants HO-ASDM and SIRR ($p < 0.05$). Between HO-ASDM and SIRR a significant difference ($p < 0.01$) can be found as well.

Trial 2 The measurement of the Minster in York was by far the most reverberant one with a $T_{30} \approx 7.9\,\mathrm{s}$ and showed the most dominant artefacts. In this trial as well, the 4DE performed significantly as the best ($p < 0.01$) compared to the other upmixing variants, which were assessed insignificantly equal among each other by the subjects. This trial indicates that the stronger the artefacts are, the bigger the differences between the 4DE and the other variants become.

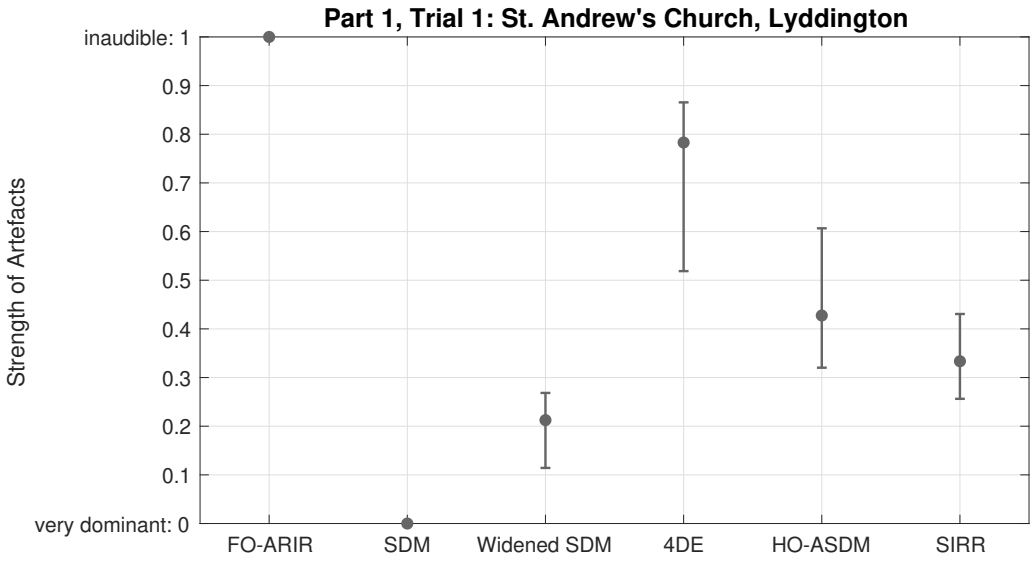

Figure 6 – Median and CI of the different stimuli presented in Part 1, Trial 1.

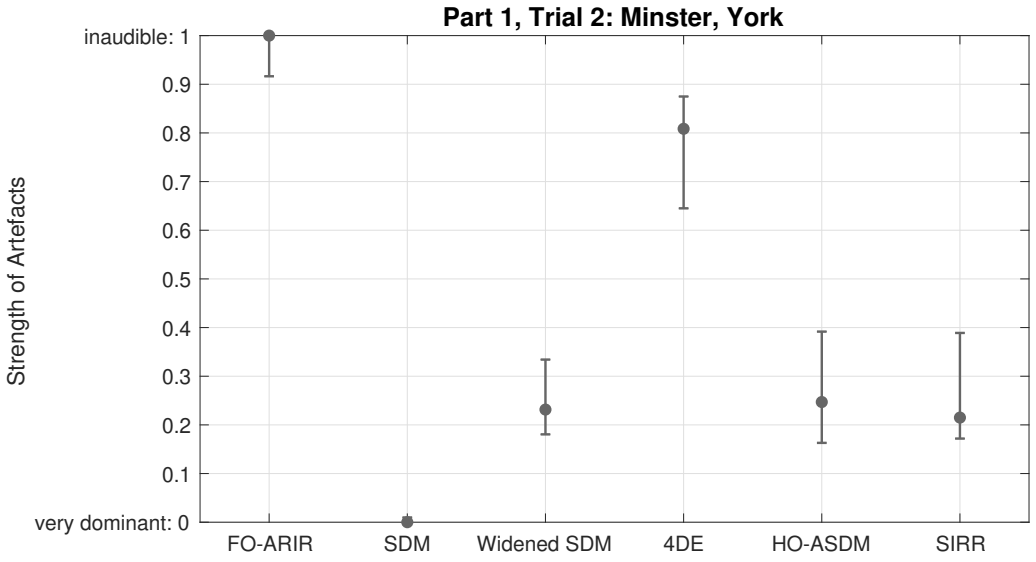

Figure 7 – Median and CI of the different stimuli presented in Part 1, Trial 2.

Trial 3 The CUBE at the institute is a very dry room ($T_{30} \approx 1\,\mathrm{s}$) and the artefacts are rather quiet. In trial 3 the trend seen in trial 1 and 2 appears as well. The 4DE is the highest rated algorithm and all groups of stimuli are significant different to each other.

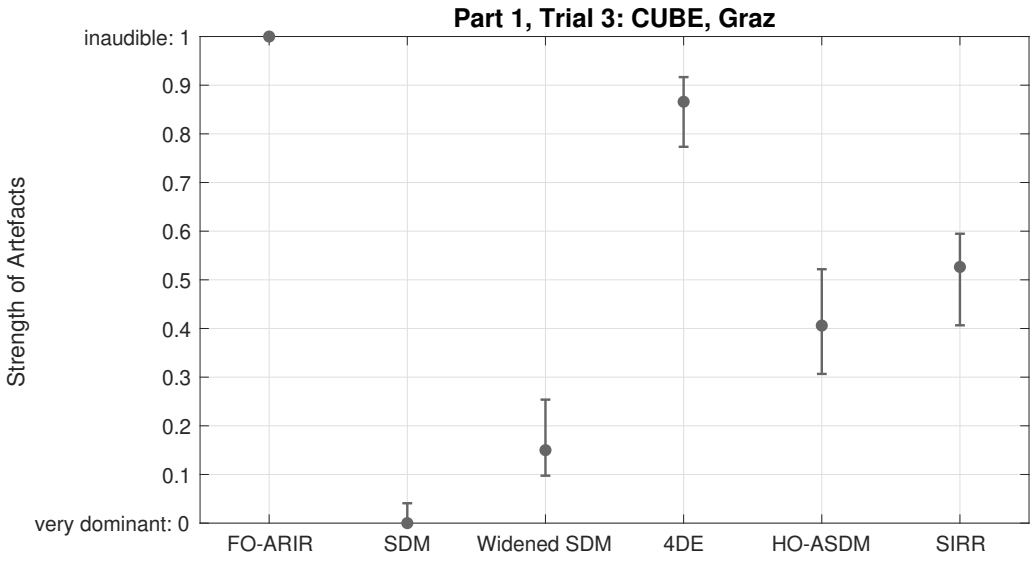

Figure 8 – Median and CI of the different stimuli presented in Part 1, Trial 3.

4.3.2 Part II: Investigation of Differences in Timbre

The preservation of the original timbre of the FO-ARIR is a crucial quality feature of an upmixing algorithm. In the first two trials of part II the performance of the SDM with Ambisonic Widening is significantly poor. Compared to part I, the opposite effect in the comparison of SDM and SDM with Widening is noticeable. The masking of the artefacts is bound to a trade-off to the coloration. This variant has to be rejected, because both of these qualities are essential for a good upmix.

Trial 1 Regarding the timbre of the ARIRs, 4DE, and HO-ASDM have the highest median values and no significant difference in the evaluations, but are significantly better rated than SIRR ($p < 0.001$). The 4DE-algorithm is even significantly better rated than the classic SDM ($p < 0.01$).

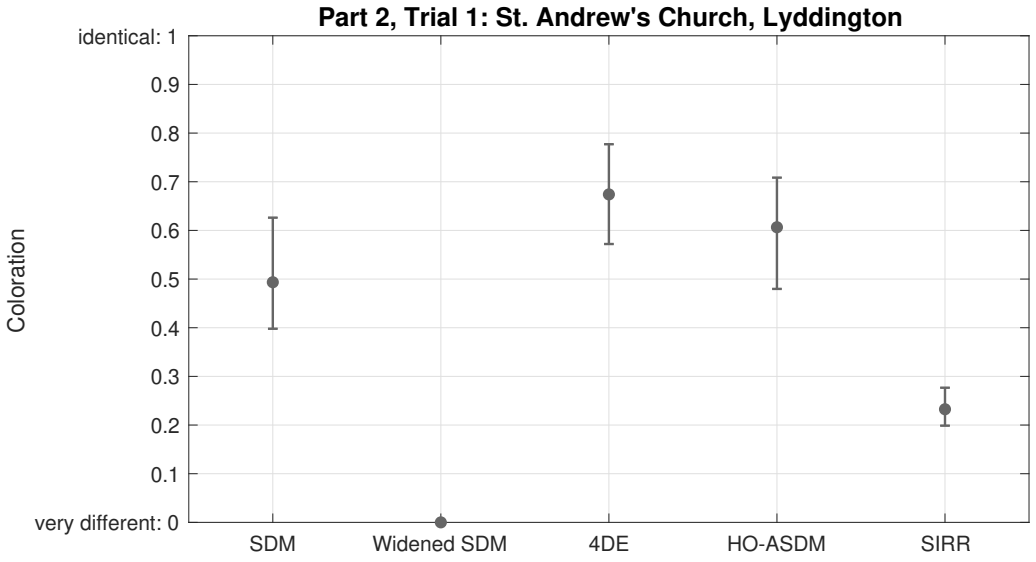

Figure 9 – Median and CI of the different stimuli presented in Part 2, Trial 1.

Trial 2 In a highly reverberant scenario, SDM, 4DE, and HO-ASDM perform equally good, again significantly better than SIRR and SDM with Ambisonic Widening. This trend is in accordance with trial 1. In this trial, there were no significant differences between SDM, 4DE, and HO-ASDM.

Trial 3 In trial 3, the stimulus of the SDM with Widening variant of the St. Paul's Concert Hall ($T_{30} \approx 1.5\,\mathrm{s}$) is exchanged with a fourth-order measurement of the concert hall. The incentive was becoming information, on how large the differences in coloration are perceived by the subjects in relation to the differences, produced by the measurement in higher-order. The result is very clear, the fourth-order measurement has a very different timbre than the reference and the upmixed signals. The stimuli can be divided into two groups, that were very significantly different ($p < 0.01$): the upmixing variants in one group and the 4o-ARIR as only element in the other group.

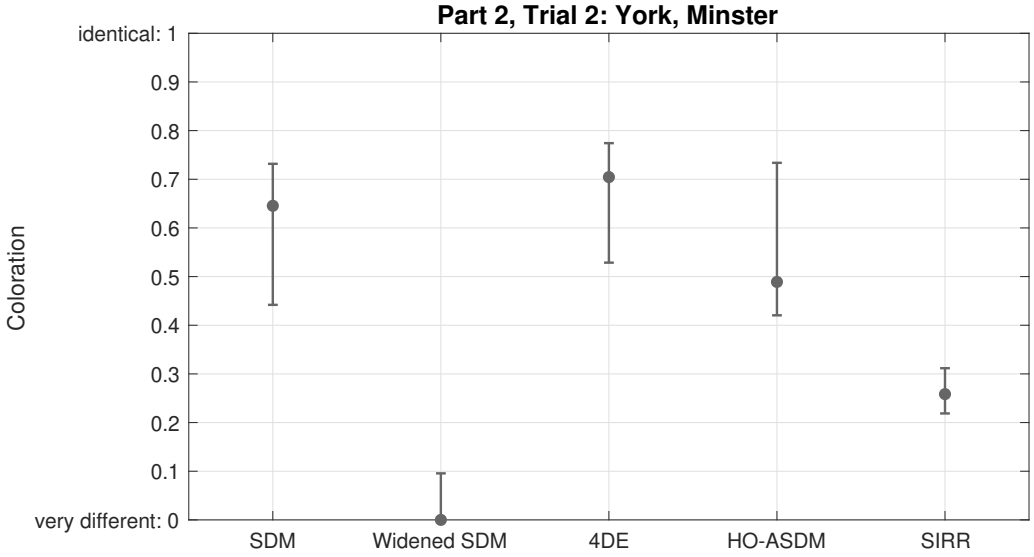

Figure 10 – Median and CI of the different stimuli presented in Part 2, Trial 2.

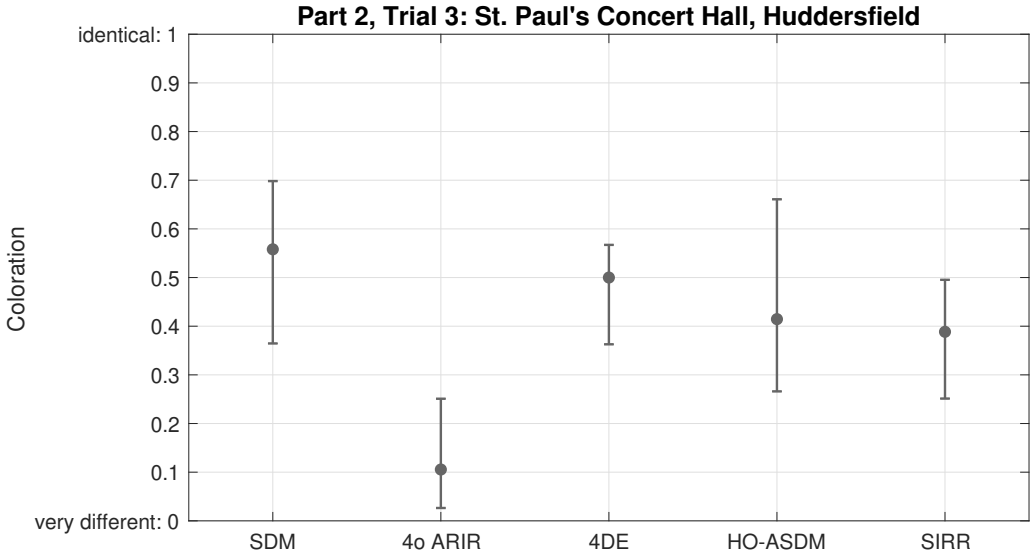

Figure 11 – Median and CI of the different stimuli presented in Part 2, Trial 3.

5 Conclusion and outlook

In this project thesis, the development and evaluation of the 4DE-algorithm is presented. The objective of this work, clearing the FO-ARIRs from artefacts could be achieved in a very successful manner. In part I of the listening tests with experienced subjects, the 4DE-algorithm is rated significantly better than the most used upmixing algorithms that are currently available, as shown in Section 4.3.
Regarding the preservation of the original coloration, tested in part II, the results were comparable to the other algorithms. Trial 3 in Part II showed, that the upmixed HO-ARIRs are more similar to the first-order measurement than the fourth-order ARIR. It is a difficult question to answer, which measurement enables to simulate the more realistic sounding reverberation on a playback-setup compared to the actual acoustic experience in the real room.
Since the listening experiment was conducted on headphones, no statements about the spatial features of the 4DE-Upmix can be made. However informal listening tests of the author and colleagues on a loudspeaker setup in an acoustically treated studio indicate, that there will be achieved good results as well. A listening test in the near future is planned to support these hints with valid data.

Still some improvements to the 4DE-algorithm can be made: An unwhitening procedure, that leaves the first-order signal untouched, promises to accomplish a better preservation of the original timbre. The impact of the technique of the DOA estimation, the number of de- and encoded directions, the treatment of the early reflections and the late reverberations leave room for further investigation and research. Regarding the fundamentals, the precise cause of the artefacts is still to be examined.

List of Figures

References

[Dep19] T. Deppisch, "Multi-direction analysis in ambisonics," Thesis, Institut für Elektronische Musik und Akustik, Kunstuni Graz, Technical University Graz, Graz, A, 2019.

[FZ16] M. Frank and F. Zotter, "Spatial impression and directional resolution in the reproduction of reverberation," *Fortschritte der Akustik - DEGA*, 2016.

[FZ17] ——, "Exploring the perceptual sweet area in ambisonics," in *Audio Engineering Society Convention 142*, May 2017.

[Mal99] D. G. Malham, "Higher order ambisonic systems for the spatialisation of sound," *International Conference on Mathematics and Computing*, 1999.

[Mül19] K. Müller, "Variable-perspective rendering of virtual acoustic environments based on distributed first-order room impulse responses," Thesis, Institut für Elektronische Musik und Akustik, Kunstuni Graz, Technical University Graz, Graz, A, 2019.

[NZDS11] C. Nachbar, F. Zotter, E. Deleflie, and A. Sontacchi, "ambix - a suggested ambisonics format," *Ambisonic Symposium 2011, Lexington, KY*, 2011.

[PML04] V. Pulkki, J. Merimaa, and T. Lokki, "Reproduction of reverberation with spatial impulse response rendering," *Journal of the Audio Engineering Society*, vol. 61, no. 1/2, pp. 17–28, May 2004.

[SH02] N. J. A. Sloane and R. H. Hardin, "Spherical designs," 2002, accessed 07-January-2021. [Online]. Available: http://neilsloane.com/sphdesigns/

[SP82] J. Stautner and M. Puckette, "Designing multi-channel reverberators," *Computer Music Journal*, vol. 6, no. 1, 1982.

[TPKL13] S. Tervo, J. Pätynen, A. Kuusinen, and T. Lokki, "Spatial decomposition method for room impulse responses," *Journal of the Audio Engineering Society*, vol. 61, no. 1/2, pp. 17–28, January 2013.

[WTFE20] Wikipedia, The Free Encyclopedia, "Tetrahedron," 2020, accessed 19-December-2020. [Online]. Available: https://en.wikipedia.org/wiki/Tetrahedron

[ZF19] F. Zotter and M. Frank, *Ambisonics - A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*. Springer, 2019.

[ZFKC14] F. Zotter, M. Frank, M. Kronlachner, and J.-W. Choi, "Efficient phantom source widening and diffuseness in ambisonics," *Proc. of the EAA Joint Symposium on Auralization and Ambisonics, Berlin, Germany*, January 2014.

A Index

φ	azimuth angle
ϑ	zenith angle
θ_i	one direction consistent of azimuth and zenith angle
b	number of third-octave band
K, k	maximum number of elements, number of element in a vector
L	maximum number of directions
M, m	maximum degree, number of degree of the Ambisonic signal
N, n	maximum order, number of order of the Ambisonic signal
$\tilde{N}$	maximum order of trimmed Ambisonic signal
Y_n^m	spherical harmonic
D	directivity
$\boldsymbol{I}$	pseudo-intensity vector
$\boldsymbol{h}_n^m$	one impulse response signal of particular order and degree
$\boldsymbol{y}(\theta_i)$	weights in vector notation, spherical harmonics evaluated at one direction
$\boldsymbol{S}_i$	decoded signal
$\boldsymbol{h}_N$	full Ambisonic room impulse response of size $K \times (N+1)^2$
$\tilde{\boldsymbol{h}}_N$	full upmixed Ambisonic room impulse response of size $K \times (N+1)^2$
$\boldsymbol{\chi}_N$	Ambisonic signal of size $K \times (N+1)^2$
$\boldsymbol{D}$	decoder matrix of size $K \times (N+1)^2$
$\boldsymbol{I}_N$	unity matrix of size $(N+1)^2 \times (N+1)^2$
$\mathbf{Y}_N$	weighting matrix of size $K \times (N+1)^2$
$\boldsymbol{S}$	matrix consisting of multiple decoded signals

B Probability Values

Part I, Trial 1: St. Andrew's Church, Lyddington

	SDM	Widened SDM	4DE	HO-ASDM	SIRR
FO-ARIR	0.001	<0.001	<0.001	0.001	0.001
SDM		0.002	<0.001	<0.001	0.001
Widened SDM			0.001	0.003	0.003
4DE				0.020	0.002
HO-ASDM					0.296

Table 1 – Probability values of Part I, Trial 1.

Part I, Trial 2: Minster, York

	SDM	Widened SDM	4DE	HO-ASDM	SIRR
FO-ARIR	<0.001	0.001	0.008	0.001	0.001
SDM		0.002	0.001	0.003	0.003
Widened SDM			0.001	1.081	0.988
4DE				0.002	0.002
HO-ASDM					1.081

Table 2 – Probability values of Part I, Trial 2.

Note: Values greater than 1 develop while correcting with the Bonferroni-Holm correction method.

Part I, Trial 3: CUBE, Graz

	SDM	Widened SDM	4DE	HO-ASDM	SIRR
FO-ARIR	<0.001	0.001	0.006	0.001	0.001
SDM		0.006	0.001	0.003	0.002
Widened SDM			0.002	0.009	0.006
4DE				0.001	0.001
HO-ASDM					0.036

Table 3 – Probability values of Part I, Trial 3.

Part II, Trial 1: St. Andrew's Church, Lyddington

	Widened SDM	4DE	HO-ASDM	SIRR
SDM	<0.001	0.005	0.0104	<0.001
Widened SDM		<0.001	<0.001	<0.001
4DE			0.0351	<0.001
HO-ASDM				<0.001

Table 4 – Probability values of Part II, Trial 1.

Part II, Trial 2: Minster, York

	Widened SDM	4DE	HO-ASDM	SIRR
SDM	0.002	0.026	0.026	0.004
Widened SDM		0.001	0.001	0.001
4DE			0.026	0.006
HO-ASDM				0.007

Table 5 – Probability values of Part II, Trial2.

Part II, Trial 3: St. Paul's Concert Hall, Huddersfield

	HO-ARIR	4DE	HO-ASDM	SIRR
SDM	0.007	0.078	0.047	0.015
HO-ARIR		0.005	<0.001	0.005
4DE			0.081	0.015
HO-ASDM				0.081

Table 6 – Probability values of Part II, Trial 3.