DOI: 10.22032/dbt.39961
DOI (proceedings): 10.22032/dbt.39936

Verband Deutscher Tonmeister e.V.

Full Reviewed Paper at ICSA 2019

Presented* by VDT.

Switched Spatial Impulse Response Convolution as an Ambisonic Distance-Panning Function

Patrick Cairns[1], Dr David Moore[2]

[1] *Glasgow Caledonian University, Glasgow, Scotland, Email: patrickcairns1991@gmail.com*

[2] *Glasgow Caledonian University, Glasgow, Scotland, Email: j.d.moore@gcu.ac.uk*

Abstract

Ambisonics offers a robust and effective approach to the recording, processing and delivery of Spatial Audio. The Ambisonic system is often considered to provide a perceptually and computationally advantageous Spatial Audio experience in comparison to typical Binaural systems. This is true even when an end-step Binaural render is required, as is typical in Virtual or Augmented Reality systems which naturally imply audio delivery via headphones.

Standard Ambisonic processing allows for the rotation of a sound field around an origin position. There is not, however, a strongly established means of modulating the radial distance of a virtual sound source from the origin.

This paper presents a potential solution to an Ambisonic distance-panning function for both static and dynamic virtual sources in the form of a FOA (First Order Ambisonics) Switched-SIR (Spatial Impulse Response) Convolution Reverberator. This includes a presentation of the framework for such a function, and an analysis of audio rendered using prototype scripts.

1. Introduction

Two prominent approaches to Spatial Audio have become popular in applications: Binaural Audio, and Ambisonics (though less used approaches such as Wave Field Synthesis are also available) [1], [2].

Ambisonics is generally considered to be an accurate and perceptually satisfying depiction of Spatial Sound, even when considered in systems where Binaural Audio rendering is required at the end step for headphone delivery [2], [3], [4], [5], [6], [7]. The principles of Ambisonics allow for an optimised Spatial Audio processing medium which is being enthusiastically adopted by the cutting edge of the audio industry.

Typical Ambisonic processing only allows for the rotation of a 3-dimensional sound field around a centre point, providing no control over sound source distance from this origin [8].

Though a degree of research has been undertaken to provide control over the distance parameter in Ambisonic processing there remains no uniform method of accomplishing this function. Current designs largely fall into one of three categories:

1. Systems which that are only typically appropriate for modelling the free field (Wave Field Synthesis, amplitude attenuation) [9], [10], [11].
2. Systems which allow modulation of listener position within a sound-field but not the modulation of the radial distance between a virtual source and sound field centre (Virtual Loudspeaker approach) [5], [6].
3. Reverberators which are only appropriate for accurately modelling regular rectangular rooms or diffuse fields [12], only model static sound sources [13], or are only viable up to a small order of reflections in real-time application (Geometric Simulation) [14], [15], [16].

It can be clearly seen that such designs, though innovative and useful, do not meet the criteria required of an Ambisonic distance panning function for modern Spatial Audio applications: The ability to accurately place or emulate the placement of sound sources in a sound field within a complex acoustic environment, the ability to render dynamic audio over varying distance, and viability in real-time application for dynamic systems.

This paper offers a solution to an Ambisonic distance-panning function that meets these criteria in the form of Switched SIR (Spatial Impulse Response) Convolution. In this system SIR sets describing a range of discrete distances for a specified acoustic environment may be convolved with a mono input signal to provide an Ambisonic Auralisation at specified distances. By 'switching' the SIR set being convolved it is possible to create Ambisonic Auralisations of dynamic sound sources moving across distance. An architecture for this solution is presented here, including a overview of the system development and design, and an assessment of audio rendered using prototype functions with relevance to the success and viability of the design (namely time-frequency analysis and listening tests).

2. Background

2.1. Ambisonics

Ambisonics is a system of full periphonic directional sound pickup, storage, processing and reproduction developed through the 1970's by Gerzon, Fellgett and Barton among others, taking influence from the earlier work of Cooper and Shiga [1], [17].

Ambisonics describes a sound field around an origin position, and with radius equal to the radial distance of a sound source from this origin position, using the spherical expansion of the wave equation in the form [18], [19], [20]:

$$p(r,\theta,\emptyset,k) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} A_n^m(k) j_n(kr) Y_n^m(\theta,\emptyset) \quad \textbf{\textit{(2.1.1)}}$$

Where $\boldsymbol{p(r,\theta,\emptyset,k)}$ is the pressure at a point in space, $\boldsymbol{k}$ is the wave number, $\boldsymbol{r}$ is the radial distance, $\boldsymbol{\theta}$ is the elevation, and $\emptyset$ is the azimuth.

The Spherical Harmonics, $\boldsymbol{Y_n^m(\theta,\emptyset)}$, and Spherical Bessel Function of the first kind, $\boldsymbol{j_n(kr)}$, describe a unit sphere in terms of functions on the surface of the sphere and radial functions respectively where $\boldsymbol{n}$ is the order and $\boldsymbol{m}$ is the degree of the Spherical Expansion.

The Spherical Harmonics are given as [18], [21]:

$$Y_n^m(\theta,\emptyset) = N_n^m P_n^m(cos\theta) e^{jm\emptyset} \quad \textbf{\textit{(2.1.2)}}$$

Where $\boldsymbol{P_n^m(cos\theta)}$ is the Legendre function, describing angle of elevation, and $\boldsymbol{e^{jm\emptyset}}$ describes the azimuth. $\boldsymbol{N_n^m}$ is a normalisation factor, typically given as the SN3D normalisation scheme [18], [21], [22]:

$$N_n^m = \sqrt{\frac{(2n+1)}{4\pi}\frac{(n-m)!}{(n+m)!}} \quad \textbf{\textit{(2.1.3)}}$$

The Spherical Bessel Function of the first kind is given as:

$$j_n(x) = (-1)^n x^n \left(\frac{1}{x}\frac{\partial}{\partial x}\right)^n \frac{\sin(x)}{x} \quad \textbf{\textit{(2.1.4)}}$$

Where in the system presented ***x=kr*** thus describing the radial functions [19], [20].

The Ambisonic Coefficients, $\boldsymbol{A_n^m(k)}$, describe the content of the sound field, and, using the free-field spherical decomposition, may be given as [18], [21]:

$$A_n^m(k) = \frac{1}{j_n(kr)}\int_0^{2\pi}\int_0^{\pi} p(r,\theta,\emptyset,k) Y_n^m(\theta,\emptyset)^* \sin(\theta)\, d\theta d\,\emptyset \quad \textbf{\textit{(2.1.4)}}$$

Each discrete order and degree of Ambisonic Coefficient corresponds to an audio channel in the Ambisonic B-Format. In First Order Ambisonics (FOA) the four B-Format audio channels describe a sound pressure in terms of an omnidirectional pressure component and plane waves along each of the three orthogonal spatial dimensions, with each component designated a discrete B-Format channel. The nomenclature for these B-Format channels has classically followed the Furse-Malham (FuMa) scheme, though more recent systems show a rising popularity in the use of the Ambisonics Exchangeable (AmbiX) scheme [23].

n	m	AmbiX ACN	FuMa Channel	A_n^m
0	0	0	W	1
1	-1	1	Y	$\sin\theta\cos\emptyset$
1	0	2	Z	$\sin\emptyset$
1	1	3	X	$\cos\theta\cos\emptyset$

Tab. 2.1.1: First Order Ambisonic Components.

2.2. Spatial Impulse Response

Impulse Response (IR) is a function describing the filtering effect of any system considering the output with respect to the input, and may be described mathematically in the convolution equation [24]:

$$y(n) = x(n) \otimes h(n) = \sum_{k=-\infty}^{\infty} x(n)h(n-k) \quad \textbf{\textit{(2.2.1)}}$$

Here ***x(n)*** is the filter input, ***y(n)*** is the filter output and ***h(n)*** denotes the impulse response.

In audio applications IR is widely used to describe the acoustic influence of a space on any sound actualised in the space. IR may be measured in acoustic spaces [25], [26], estimated using statistical approaches [27], or rendered through geometric simulation [16], [28]. IR measurement procedure considers an excitation signal as the system input and the recorded or simulated output (typically from an omni-directional microphone) as the output. The excitation signal used is required to provide an even signal across the time and frequency domain such that the acoustic properties of the room may be measured evenly. Several approaches are available for providing this excitation signal such as MLS or Sine Sweep method [25].

The Logarithmic Sine Sweep signal, as used in this paper, takes the form [25], [29]:

$$s(t) = \sin[Ke^{-\frac{t}{L}} - 1] \quad \textbf{\textit{(2.2.2)}}$$

Where ***s(t)*** is the excitation signal, ***t*** is time and ***K*** and ***L*** are given as:

$$K = \frac{\omega_1 T}{\ln\left(\frac{\omega_1}{\omega_2}\right)} \quad \textbf{\textit{(2.2.3)}}$$

$$L = \frac{T}{\ln\left(\frac{\omega_1}{\omega_2}\right)} \qquad \textbf{\textit{(2.2.4)}}$$

Where $\boldsymbol{T}$ is the duration of the sweep, and $\boldsymbol{\omega_1}$ and $\boldsymbol{\omega_2}$ are the start and end frequencies for the sweep respectively.

The impulse response is obtained for the sine sweep method by creating an inverse filter, ***f(t)***, from the excitation signal, ***s(t)***, and convolving with the room response to the excitation signal, ***r(t)***, in order to obtain the linear impulse response, ***h(t)***:

$$h(t) = r(t) \otimes f(t) \qquad \textbf{\textit{(2.2.5)}}$$

The inverse filter, ***f(t)***, is calculated by reversing the excitation signal and then applying an amplitude modulation filter of +6dB per octave. This modulation filter may be given as:

$$m(t) = \frac{\omega_1}{\omega(t)} \qquad \textbf{\textit{(2.2.6)}}$$

Where the introduced term, ***ω(t)***, is the instantaneous frequency for each sample of ***t***.

Spatial Impulse Response (SIR), also sometimes referred to as Directional Room Impulse Response (DRIR), differs in that the output of the system is the B-Format audio channels rather than the mono audio channel delivered by standard IR measurement [30], [31]. As such SIR is given in IR sets describing the difference between a mono excitation signal and each of the B-Format channels.

It should be noted that with mono-Ambisonic upmixing widely available it is a simple task to render an excitation signal as B-Format channels and obtain SIR sets describing the input-output difference for discrete Ambisonic components (B-format to B-Format rather than Mono to B-Format).

SIR may be measured using B-Format microphones, or rendered through the placement of virtual B-Format transducer arrays in geometric simulation. The obtained SIR may be applied to audio in order to render that audio in the measured acoustic environment as an Ambisonic sound field with great accuracy by convolving the SIR set with audio signal input. This process is described as Ambisonic Auralisation [30], [32].

2.3. Real-Time Convolution

Direct convolution of an impulse response and audio file is a computationally expensive method of rendering audio. The amount of data that is required to be processed in a short space of time in audio applications requires a further set of fast-convolution techniques [33].

Computational cost may be saved using Fast Fourier Transform (FFT) methods to compute the Transforms, achieve convolution through multiplication in the frequency domain, and compute the Inverse Fast Fourier Transform (IFFT) to give the convolved output [24]. This FFT method is generally based on the ‘Cooley-Tukey algorithm’, or ‘Radix-2 decimation’ [34], [35].

If the Fourier transform is given for an N-term sequence as:

$$F[k] = \sum_{n=0}^{N-1} f[n] e^{-2jnk\pi/N} \qquad \textbf{\textit{(2.3.1)}}$$

Then by declaring a variable called the ‘twiddle factor’, defined as:

$$W = e^{-2j\pi/N} \qquad \textbf{\textit{(2.3.2)}}$$

Substituting in the ‘twiddle factor’:

$$F[k] = \sum_{n=0}^{N-1} f[n] W^{nk} \qquad \textbf{\textit{(2.3.3)}}$$

Where $\boldsymbol{W}$ is constant for a fixed value of $\boldsymbol{N}$. It is then possible to take advantage of the properties of symmetry and periodicity to drastically decrease the number of computations required.

Symmetry: $W_N{}^{k[N-n]} = W_N{}^{-kn} = (W_N{}^{kn})^*$ ***(2.3.4)***

Periodicity: $W_N{}^{kn} = W_N{}^{k[n+1]} = W_N{}^{[k+N]n}$ ***(2.3.5)***

Breaking down the transform this way decreases the amount of computations required from a factor of N^2 for the Discrete Fourier Transform to a factor of NlogN [34].

Partitioned convolution algorithms break down the input signal and impulse response into blocks of samples which may be convolved in real-time using FFT algorithms [35]. The most common partitioning scheme is the Overlap-Add method, where the output of each block of convolution is summed into the system output at the relevant sample index as defined by the partitioning algorithm [33].

The overlap-add process can be outlined as [33]:

1. Partition the input signal into segments
2. Zero Pad the input blocks and impulse response to an equal and even length of FFT.
3. For each zero-padded input segment perform the FFT, Frequency Domain multiplication and IFFT.
4. For each resultant block sum into the output from the relevant sample.

3. System Overview

The Ambisonic distance-panning function offered in this paper uses SIR convolution to create Ambisonic Auralisations of a mono input signal. SIR sets, where each SIR describes a discrete distance, may be obtained through acoustic measurement or simulation. Each discrete SIR in such an obtained set describes an impulse function in terms of a spherical sound field of specified radial distance. Convolution of a mono sound source with any discrete SIR therefore renders the input as a virtual source at the relevant radial distance.

In typical overlap-add partitioned convolution algorithms a single IR is called for convolution each time a new block of the input signal is partitioned [33]. In the switched-SIR system presented here the SIR called for convolution can be varied at each partition. If we consider the instance where a sequence of SIRs are called which describe a location A and progress sequentially towards location B then the resultant Ambisonic Auralisation can be considered an emulation of a dynamic sound source.

3.1. Practical Considerations

The system developed only extends to First Order Ambisonics (FOA), largely due to the availability of equipment. It can be recognised that the design presented may indeed be extended to include higher order Ambisonic systems [13].

The system receives mono input, as this is the most common input form for Auralisation purposes [8]. It can, however, be easily seen that through various upmixing capabilities SIR sets for other input formats may be easily computed.

3.2. Previous Work

The Switched-SIR convolution design draws largely from two key systems:

1. The convolution panning system presented by Stewart and Sandler [36]. This system presents the key framework for switched-IR convolution as a panning function, and in turn takes roots in the head-tracking system developed by Reilly and McGrath which uses the same functionality with HRTFs [37].
2. The Ambisonic convolution architecture presented by Lopez-Lezanco [13] which provides the framework for the application of SIR convolution.

4. System Design

4.1. SIR Measurement

In order to obtain an SIR set which describes the distance dimension a set of real SIRs were measured in Lecture Theatre W011 at Glasgow Caledonian University.

The room was in an unoccupied state, and the noise floor was recorded at 40dB.A Dodecahedral Loudspeaker calibrated at 80dB (not ideal but more importantly within safe listening levels) was set up and used to output the excitation signal. The excitation signal used was generated in Reaper using the ReaVerb plugin to provide a 1.5s Logarithmic Sine Sweep.

A SoundField ST250 microphone was used to record the B-Format room response to the excitation signal at discrete source-receiver distances, varying from 1m to 12m at 0.5m intervals. Both source and receiver were set at 1.5m height, taking care not to vary azimuth or elevation as distance was varied. The ReaVerb plugin was then used to deconvolve the SIR on a channel-by-channel basis and SIRs were stored as .wav files.

At each discrete SIR measurement distance a B-Format recording is made of a semi-anechoic speech sample in order to provide reference material for system analysis.

4.2. Rendering Audio

A set of prototype functions were developed using MATLAB to render audio to assess the Switched-SIR distance panning design using the recorded SIR set describing discrete points along the distance dimension. A mono semi-anechoic speech sample was selected as the input signal for providing a range of static and dynamic Ambisonic Auralisations. The computer used to run these scripts features a 4GHz quad-core processor and 16GB of DDR4 RAM.

4.2.1. Static Render

The MATLAB script used to provide the static Ambisonic Auralisations consisted of a simple convolution using the Radix-2 Decimation FFT algorithm.

The mono speech sample and SIR at a manually specified distance are loaded by the script. The SIR being convolved is zero-padded to the length of the speech sample. FFTs are performed on each channel of the SIR and on the mono input signal. The mono input signal is then convolved with each channel of the SIR through multiplication in the Fourier (frequency) domain. IFFTs are then taken for each of the resultant outputs, providing the FOA B-Format output channels containing the static Ambisonic Auralisation.

This script was used to render B-Format audio at distances of 1m, 2m, 4m, 8.5m and 12m.

This process was measured as taking 0.196037s to process an excess of 2 million samples, just under 0.1 microseconds per sample, and again clearly showing that real-time application is viable with the implementation of partitioned convolution algorithms, given that at 44.1KHz we are passing 1 sample every 22.7 microseconds, this shows that block sizes of around 200 samples are a viable option for partitioning of four channels.

4.2.2. Dynamic Render

The MATLAB script used to provide the dynamic renders introduces the switched-SIR partitioning scheme in timed offline renders.

Firstly the complete SIR set describing the distance dimension and the mono speech sample are loaded by the script. The SIR set is then zero padded to equal length and FFTs are performed on each SIR. In application the pre-computing of SIR FFTs is sensible to save computing cost during online application.

The dynamic renders are specified for a dynamic sound source moving at a speed of 1m/s over 4.5m to maintain consistency for analysis. As such the input signal is partitioned at 0.5s intervals. Each partition is convolved with a discrete SIR, switching the SIR at each partition to move sequentially along the distance dimension. These convolutions are performed using the same FFT algorithm as with the static renders, and implement the overlap-add method to sum each partitioned block into the output B-Format channels.

This script was used to render dynamic sound sources travelling across distance from 1m to 5.5m, 4m to 8.5m, 7m to 11.5m, 6m to 1.5m, 9m to 4.5m, and 12m to 7.5m.

This section can be measured as taking 0.880042 seconds to complete the convolution and Inverse Fourier Transform for over 40 million samples, indicating once again that real-time application of such a system is easily viable with real-time partitioned convolution techniques given that this time can be reduced using more elegant algorithms.

4.2.3. Binaural Render

In order to provide a point of comparison with typical Spatial Audio distance-panning technology equivalent-distance static and dynamic Binaural Renders were created using the mono speech sample and a typical Binaural processing tool, which allows distance modulation through simple amplitude attenuation and the computation of reflections up to the 3rd order.

5. System Analysis

The success and viability of the developed Switched-SIR Ambisonic distance panning function for rendering static and dynamic B-Format audio was assessed through Time-Frequency Analysis and Listening Tests.

5.1. Time-Frequency Analysis

5.1.1. Static Ambisonic Renders

Analysis of static Ambisonic renders was conducted in order to examine similarity to the reference recordings made during the SIR measurement process.

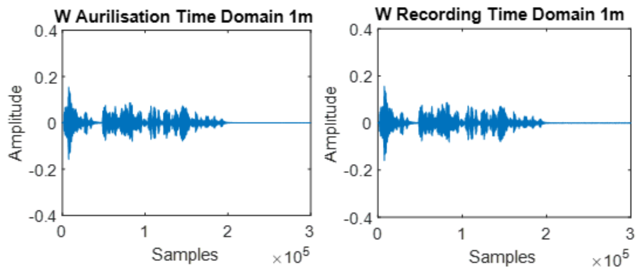

Fig. 5.1.1.1: Auralised Vs Recorded B-Format W channel at 1m.

Observing in the time domain it is evident how accurate an Auralisation SIR Convolution provides even when using FFT algorithms.

Peak-matching between the Auralisation and recordings consistently showed only miniscule differences between iterations of each audio channel.

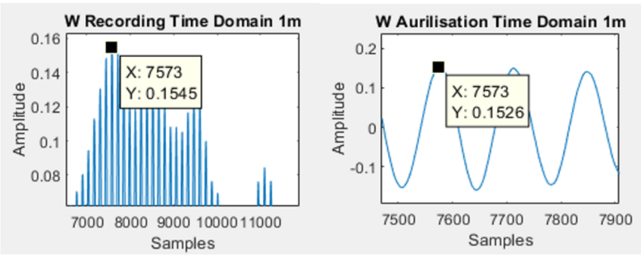

Fig. 5.1.1.2: Peak-matching between recordings and static Auralisation at 1m.

Spectral comparison between the static Auralisations and recording did however show certain inconsistencies between the two.

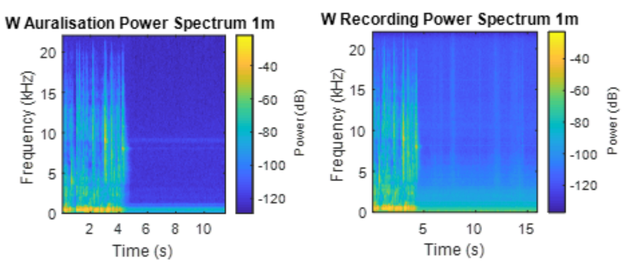

Fig. 5.1.1.3: Power Spectrum of recordings and static Auralisation at 1m.

Though extremely accurate the Auralisations notably contain more 'defined' spectral qualities (showing a lessening of the spectral smearing associated with reverberation) and contained less late high-frequency energy, presumably due to the length of the SIRs.

The smearing at around 8-9kHz and mains hum apparent in the Auralisations are simply the effects of non-ideal SIR measurement in a busy city centre campus.

5.1.2. Dynamic Ambisonic Renders

As no options were available for providing dynamic Ambisonic recordings of the mono speech sample the dynamic Ambisonic Auralisations are rendered as Binaural Stereo and evaluated with comparison to the dynamic Binaural renders created using the typical Binaural processing tool.

In the time domain the Ambisonic Auralisation can be seen as an accurate render when viewed side-by-side to the Binaural-only processing equivalent, indicating that the Switched-SIR method does indeed provide a valid means of rendering dynamic sound sources. The Ambisonic Auralisation can also be seen as providing more natural tails than the Binaural processing where the reverberant levels quickly fall due to the low order of computed reflections.

This is also visible in the waterfalls plots.

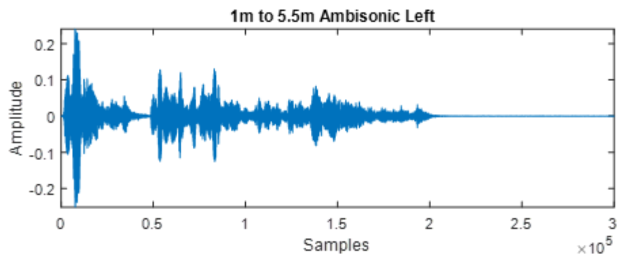

Fig. 5.1.2.1: Dynamic Ambisonic Auralisation from 1m to 5.5m in time domain, left Binaural channel

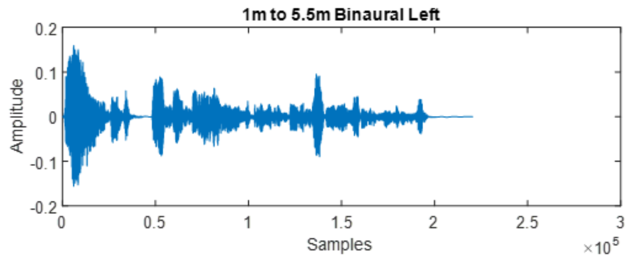

Fig. 5.1.2.2: Dynamic Binaural Render from 1m to 5.5m in time domain, left Binaural channel.

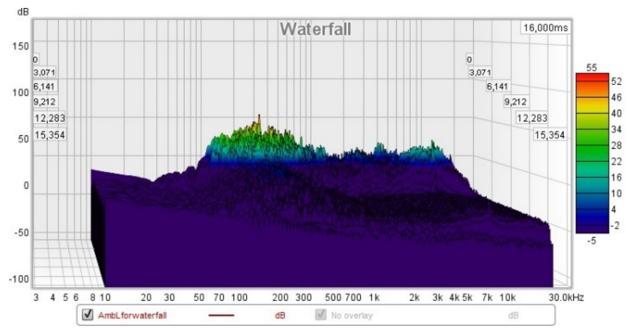

Fig. 5.1.2.3: Left Channel Waterfall of Binaural End-Step Render of Dynamic Ambisonic Auralisation moving from 7m to 11.5m.

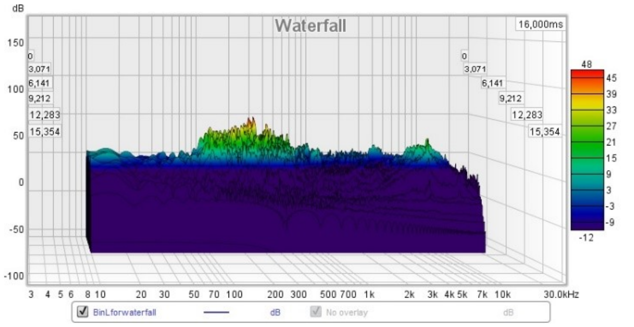

Fig. 5.1.2.4: Left Channel Waterfall of Binaural process Render of Dynamic Sound Source moving from 7m to 11.5m.

5.2. Listening Tests

Listening tests were conducted to assess the ability of the developed system to accurately deliver auditory distance cues for both static and dynamic renders.

As a full Ambisonic playback system was impractical, and again end-step Binaural renders are a typical component in Spatial Audio systems, audio for these listening tests were rendered as Binaural Stereo. This allows an assessment of the switched SIR Ambisonic distance panning function as a processing option for Binaural delivery in contrast to typical Binaural distance processing.

As absolute source-receiver distance is notoriously inaccurately perceived using only auditory cues where a user has no other indication to their environment a more meaningful measure of distance perception was utilised: relative distance between two sound sources.

5.2.1. Static Renders

13 experienced listeners were asked to indicate the absolute perceived distance of a static sound source on a continuous numerical scale. Static renders at 1m, 2m, 4m, 8.5m and 12m were played via headphones in random order. Using the first indicated perceived distance as a ‘reference’ measurement it was then possible to extract information on the percentage error of accuracy in relative distances perceived between sound sources rendered at varying distance.

This process was conducted for static Ambisonic Auralisation, static Ambisonic Recordings and static Binaural-only processing, and results were subjected to a one-way Analysis of Variance (ANOVA).

Static Ambisonic Recordings indicated that the accuracy of perceived relative distance does indeed vary significantly over a range of distances. The significant outlier can be clearly identified as the 3m discrepancy, this is likely due to the nature of percentage errors, and indicates some testing redesign may be in order.

ANOVA of results from the Static Ambisonic Auralisations indicated that relative distance can indeed be perceived with consistent accuracy over varying distances, though again the largest percentage error can be seen in smaller discrepancies. Otherwise results exhibited a similarity to the results from the Ambisonic recordings. The increased accuracy of the Ambisonic Auralisation compared to Ambisonic recordings can presumably be attributed to the reduced late reverberant energy, thus reducing the natural smearing effect of reverberation on sound source localisation [38], [39].

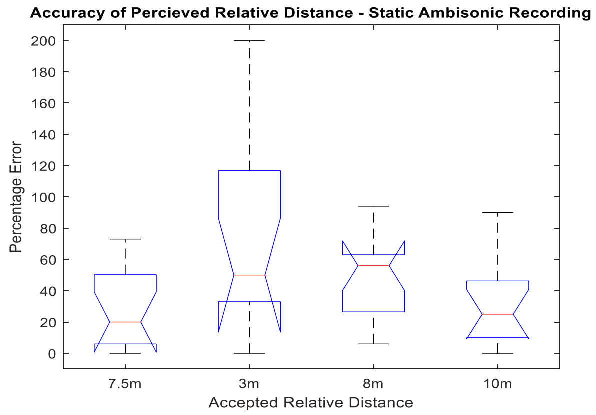

Fig. 5.2.1.1: Notch Graph of results from Static Ambisonic Recording listening tests.

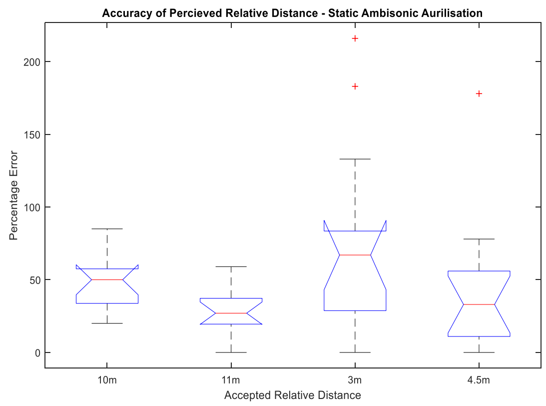

Fig. 5.2.1.2: Notch Graph of results from Static Ambisonic Auralisation listening tests.

The Binaural-processing render listening test results showed another significant outlier at 3m, indicating relative distance is not perceived consistently as distance varies, but did however otherwise show greater accuracy than expected from natural listening conditions. This could again suggest that the lack of natural reverberation in the rendering provides unnaturally accurate localisation [38], [39].

5.2.2. Dynamic Renders

13 experienced listeners were asked to indicate the start and stop distance of a dynamic sound source played over headphones on a continuous numerical scale. Sound sources moved over distances of 1m to 5.5m, 4m to 8.5m, 7m to 11.5m, 6m to 1.5m, 9m to 4.5m, and 12m to 7.5m. This test was undertaken under the same form for the audio rendered from Binaural-only processing and from the Switched SIR Ambisonic Distance panning function presented in this paper.

One-way ANOVA of results showed that neither approach yielded consistent perception of relative distance over varying distances. The results were, however, extremely similar, indicating the Switched SIR system performs comparably to the Binaural system in this regard. The inconsistency is considered as possibly due to the overall poor performance of the human auditory system without accompanying perceptual cues, and that these results mirror these expectations from natural listening conditions [38] [39].

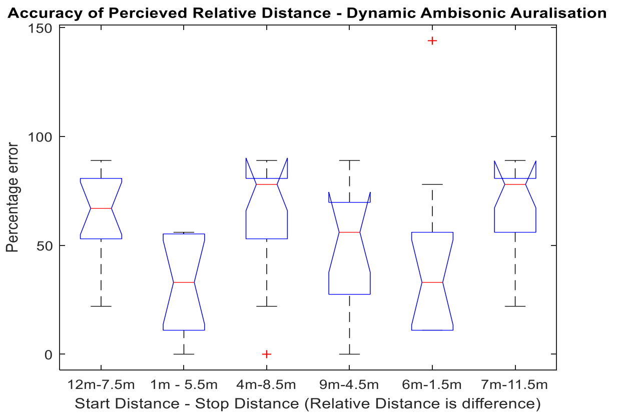

Fig. 5.2.2.1: Notch Graph of results from Dynamic Ambisonic Auralisation listening tests.

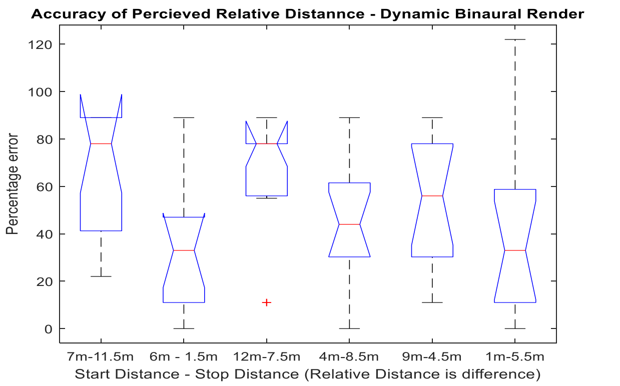

Fig. 5.2.2.2: Notch Graph of results from Dynamic Binaural render listening tests.

6. Conclusion

Time-frequency analysis of static renders confirms that Ambisonic Auralisation is an effective method of rendering a mono input as an Ambisonic sound source across distance. Results would indicate that the richer reverberant field spectral content of the Ambisonic Auralisations when compared to typical Binaural processing would indicate that Ambisonics is indeed a superior processing medium with a closer resemblance to natural listening conditions, though comparison to Binaural Impulse Response Auralisation would be required to confirm this.

Timed renders using prototype scripts confirm that the Ambisonic convolution reverb architecture presented by Lopez-Lezanco [13] is possible to apply in real-time using existing FFT and partitioned convolution algorithms.

Time-Frequency Analysis of the developed Switched SIR Ambisonic distance panning function did effectively render B-Format audio dynamically across varying distance. Such renders exhibited the same richer reverberant field and spectral content when compared to typical Binaural processing as was apparent in the static renders. This confirms that the developed system is an appropriate solution to distance-panning in the Ambisonic medium.

Timed renders of dynamic Ambisonic Auralisations using Switched SIR partitioning, FFT and overlap-add algorithms was again suitable for real-time application using existing techniques.

Listening test results were largely inconclusive, though did suggest that the Ambisonic processing more accurately depicts natural listening than Binaural rendering, as was noted several times in a ‘Free Comment’ option given to participants, though this requires further investigation.

The spatial resolution used in this prototype development and analysis was only that of 0.5m for dynamic sound sources. It was noted that little information on Just Noticeable Difference in auditory distance perception is available, and investigation into this would allow for some optimisation in spatial resolution in such systems.

It is also noted that in typical Ambisonic processing rotation of a sound field implies rotation of the acoustic environment. For many applications this is inappropriate, and as such it is proposed that the same functions used in providing Switched SIR Convolution as an Ambisonic Distance Panning function may be extended to include Azimuth and Elevation functions.

7. References

[1] F. Rumsey, Spatial Audio, Oxford: Focal Press, 2001.

[2] W. Zhang, P. N. Samarasinghe, H. C. and T. D. Abhayapala, “Surround by Sound: A Review of Spatial Audio Recording and Reproduction,” *Applied Sciences,* vol. 7, no. 532, p. 1, 2017.

[3] E. Benjamin, R. Lee and A. Heller, “Why Ambisonics Does Work,” in *AES 129th Convention*, San Francisco, 2010.

[4] A. Tarzan, M. Alunno and P. Bientinesi, “Assessment of Sound Spatialisation Algorithms for Sonic Rendering with Headsets,” Aachen university, Aachen, 2017.

[5] G. Kearney, M. R. H. Gorzel and F. Boland, “Distance Perception in Interactive Virtual Acoustic Environments Using First and Higher Order Ambisonic Sound Fields,” *Acta Acustica United Wih Acustica,* vol. 98, no. 1, pp. 61-67, 2012.

[6] A. McKeag and D. McGrath, “Sound Field Format to Binaural Decoder with Head-Tracking,” in *AES 6th Australian Regional Convention*, Melbourne, 1996.

[7] M. Noisternig, A. Sontacchi, T. Musil and R. Holdrich, “A 3D Ambisonic Based Binaural Sound Reproduction System,” in *AES 24th International Conference on Multichannel Audio*, Banff, 2003.

[8] T. Nishiyama, T. Nagata, S. Ogata, T. Hasue and M. Kashiwagi, “Recording and mixing techniques for Ambisonic sound reproduction,” in *AES International Conference on Spatial Reproduction*, Tokyo, 2018.

[9] R. Penha, "Distance Encoding in Ambisonics Using Three Angular Coordinates," in *Sound and Music Computing Conference*, Berlin, 2008.

[10] D. Menzies and M. Al-Akaidi, "Ambisonic Synthesis of Complex Sources," *Journal of the Audio Engineering Society,* vol. 55, no. 10, pp. 864-875, 2007.

[11] A. Sontacchi and R. Holdrich, "Distance Coding in 3d Sound Fields," in *AES 21st Conference*, St. Peterburg, 2002.

[12] B. Wiggins and M. Dring, "AmbiFreeverb 2 - Development of a 3D Ambisonic Reverb with Spatial Warping and Variable Scattering," in *AES Conference on Sound Field Control*, Guildford, 2016.

[13] Lopez-Lezcano, "An Architecture for Reverberation in High Order Ambisonics," in *AES 137th Convention*, Los Angeles, 2014.

[14] A. Oliveira, G. Campos, P. Dia, D. Murphy, J. Viera, C. Mendonca and J. Santos, "Real-Time Dynamic Image Source Implementation for Auralisation," in *16th International Conference on Digital Audio Effects*, Maynooth, 2013.

[15] S. McGovern, "The Image-Source Reverberation Model in an N-Dimensional Space," in *14th International Conference on Digital Audio Effects*, Paris, 2011.

[16] A. Krokstad, S. Dtrom and S. Sorsdal, "Calculating the Acoustical Room Response by Using the use of a Ray Tracing Technique," *Journal of Sound Vibrations,* vol. 8, no. 1, pp. 118-125, 1968.

[17] M. A. Gerzon, "Periphony: With Height Sound Reproduction," *Journal of the Audio Engineering Society,* vol. 21, no. 1, pp. 2-10, 1973.

[18] M. A. Poletti, "Three-Dimensional Surround Sound Systems Based on Spherical Harmonics," *Journal of the Audio Engineering Society,* vol. 53, no. 11, pp. 1004-1025, 2005.

[19] L. L. Beranek and T. J. Mellow, Acoustic Sound Fields and Transducers, London: Academic Press, 2012.

[20] J. Ahrens, Analytic Methods of Sound Field Synthesis, Berlin: Springer, 2012.

[21] J.-M. Batke, "The B-Format Microphone Revised," in *Ambisonics Symposium*, Graz, 2009.

[22] C. Nachbar, F. Zotter, E. Deleflie and A. Sontacchi, "Ambix - A Suggested Ambisonic Format," in *Ambisonics Symposium*, Lexington, 2011.

[23] D. Malham, "Space in Music - Music in Space: Masters Thesis," Uniersity of York, York, 2003.

[24] A. V. Oppenheim and R. W. Schfer, Discrete Time Signal Processing, London: Prentice-Hall, 1999.

[25] G.-B. Stan, J.-J. Ebrechts and D. Archambeau, "Comparison of Different Impulse Response Measurement Techniques," *Journal of the Audio Engineering Society,* vol. 50, no. 4, pp. 249-262, 2002.

[26] M. Kleiner, B.-I. D. And and P. Svensson, "Auralization - An Overview," *Journal of the Audio Engineering Society,* vol. 41, no. 11, pp. 861-875, 1993.

[27] J.-M. Jot, L. Cerveau and O. Warusfel, "Analysis and Synthesis of Room Reverberation Based on a Statistical Time-Frequency Model," in *AES 103rd Convention*, New York, 1997.

[28] N. Tsingos, "Pre-computing geometry-based reverberation effects for games," in *AES 35th International Conference*, London, 2009.

[29] A. Farina, "Simultaneous Measurement of Impulse Response and Distortion with a SIne Swept technique," in *AES 108th Convention*, Paris, 2000.

[30] J.-D. Polack and F. L. Figueiredo, "Room Acoustic Auralization with Ambisonics," in *Societe Francaise D'Acoustique*, Nantes, 2012.

[31] A. Perez-Lopez and J. De Munke, "Ambisonic Directional Room Impulse Response as a new Convention of the Spatially Oriented Format for Acoustics," in *AES 144th Convewntion*, Milan, 2018.

[32] J. J. Embrechts, "Review on the Application of Directional Impulse Responses in Room Acoustics," in *Congrès français d'acoustique*, Le Mans, 2016.

[33] U. Zolzer, Digital Audio Effects, Chichester: John WIley and Sons, 2011.

[34] D. Lyon, "The Discrete Fourier Transform, Part 2: Radix 2 FFT," *Journal of Object Technology,* vol. 8, no. 5, pp. 21-33, 2009.

[35] F. Wefers, "Partitioned Convolution Algorithms for Real-Time Auralization: PHD Thesis," Logos Verlag, Berlin, 2014.

[36] R. Stewart and M. Sandler, "Real-time Panning Convolution Reverb," in *AES 12rd Convention*, New York, 2007.

[37] A. Reilly and D. McGrath, "Real-Time Auralization with Head Tracking," in *AES 5th Australian Regional Convention*, Sydney, 1995.

[38] J. Blauert, Spatial Hearing: The Psychophysics of Human Sound Localization (Revised Edition), London: MIT Press, 1999.

[39] S. H. Nielsen, "Auditory Distance Perception in Different Room," in *AES 92nd Convention*, Vienna, 1992.

