Higher-Order Stereophony

Jacob Hollebon and Filippo Maria Fazi

***Abstract*—This work introduces a new theory for spatial audio recording and reproduction named Higher Order Stereophony. Through the use of the Taylor expansion, the technique accurately reproduces a sound field across a line that is orientated as the interaural axis of a listener, to attempt to recreate a set of desired binaural signals. The technique utilises loudspeaker amplitude panning, and is shown to encompass in its framework traditional Stereophony approaches such as the stereo sine law. Therefore, the technique expands Stereophony to higher orders and more loudspeakers, leading to a greater frequency range of accurate reproduction, in a similar manner to Higher Order Ambisonics. Higher Order Stereophony is shown to exhibit many similarities to Higher Order Ambisonics, and decoders to transition between the different sound field representations are derived. Higher Order Stereophony is also re-derived through a mode matching approach using a subset of spherical harmonics, those with degree index equal to zero only. The theoretical results are then validated through experimental measurements using various microphone arrays, considering the reproduced sound field across a single line and the reproduced spherical harmonic coefficients of the sound field.**

***Index Terms*—Spatial Audio, Panning, Stereophony, Higher Order Ambisonics (HOA).**

I. Introduction

THE goal of a spatial audio system is to reproduce the acoustic illusion of virtual acoustic scenes to a listener. In its most classic case, this takes the form of a virtual sound source positioned somewhere in 3D space about the listener. Most commonly, loudspeaker amplitude panning is utilised to create the illusion of a virtual source through the phenomenon of summing localisation [1], [2]. Interestingly, some panning approaches such as Higher-Order Ambisonics (HOA) can also reproduce the physical properties of the target sound field over a region of space, through calculating loudspeaker gains by representing the sound field in terms of an orthogonal basis such as the spherical harmonics. The HOA mode matching approach leads to sound field reproduction over a circular or spherical region about an expansion point in 2D and 3D, respectively [3], [4].

The Taylor expansion is a method to represent an analytical function by means of its derivatives evaluated about an expansion point. However, the Taylor expansion has found little use in sound field reproduction literature, mainly with the work by Dickins at the turn of the century [5], [6]. Dickins compared the multivariable (3D) variant of the Taylor expansion to the spherical harmonic expansion of a sound field. When restricting to physical sound fields that satisfy the wave equation and considering that both descriptions contain the same amount of information, the Taylor expansion was found to be over-specified resulting in a greater number of terms on truncation to the N-th order when compared to the spherical harmonic expansion. Interestingly, this is true only when $n \geq 2$, because the zeroth and first order for both expansions contain equal number of terms. It was therefore concluded that the spherical harmonic expansion was more compact and thus the most convenient sound field descriptor when considering 3D sound fields. This has likely led to many preferring to use the spherical harmonic expansion, for example as noted in the seminal work by Poletti [2]. Some work has utilised the Taylor expansion to first order only for analysis of spatial audio reproduction systems [7]–[9], and amplitude panning with listener head-tracking [10], [11].

The main contribution of this work is the introduction of a new approach for spatial audio reproduction titled Higher-Order Stereophony (HOS). HOS reproduces the sound field accurately across a line which is designed to align with the listener's interaural axis, through the use of the single variable (1D) Taylor expansion. This creates an efficient sound field reproduction approach, which only aims to recreate the sound field correctly at the position of the listener's ears, ideally leading to the reproduction of the desired binaural signals. This simplification leads to a significant reduction in the number of loudspeakers required and their positioning compared to other sound field reproduction approaches, as N-th order stereo requires only $(N + 1)$ loudspeakers. The classic stereo sine law is derived using this framework, and the new approach is shown to generalise classic stereo to higher orders and generalised loudspeaker arrays, in a similar manner as HOA generalises First Order Ambisonics. HOS is verified as a sound field reproduction technique across a line through measurements of the reproduced sound field using a linear microphone array.

A secondary contribution of the article is the demonstration of a fundamental relationship between HOA and HOS. Decoders are derived to transition between the two sound field representations, ensuring all existing HOA content can be reproduced using simpler and smaller loudspeaker arrays through the HOS approach. An alternative derivation for HOS is also presented through mode matching using a subset of spherical harmonics. These results are also verified using measurements utilising a spherical microphone array.

The article is arranged as follows. First, the theory of the technique is presented, resulting in a set of order matching equations (analogous to mode matching) that define the necessary HOS loudspeaker gains for any given loudspeaker array. Next, the classic stereo sine law is derived through the new HOS framework and it is demonstrated that the technique

This paper was produced by the IEEE Publication Technology Group. They are in Piscataway, NJ.

Manuscript received April 19, 2021; revised August 16, 2021.

generalises the traditional stereo technique to higher orders and any number of loudspeakers. Next, the link between HOA and HOS is explored and decoders are derived to transition between the two representations. Finally, experiments using various microphone arrays are used to validate the approach.

II. The Taylor Expansion

A. The Single Variable Taylor Expansion

The Taylor expansion expresses a well-behaved function about an expansion point as a infinite summation of its derivatives evaluated at the expansion point. In 1D, if $p(x)$ is an infinitely differentiable function at a point x_0, then [12]

$$p(x) = \sum_{n=0}^{\infty} \frac{(x-x_0)^n}{n!} \frac{d^n p(x_0)}{dx^n}. \tag{1}$$

The n-th order term depends on the n-th derivative. Practically, the infinite summation must be truncated to a finite order N introducing an error into the representation. In this case, increasing the order N results in a better approximation for larger arguments of $(x - x_0)$ and thus the approximation's accuracy further away from the expansion point x_0.

B. Expansion Of A Plane Wave sound field

The goal of a 3D audio reproduction system is to reproduce a given sound field by recreating the correct binaural signals at the listener's ears. The HOS approach is to suggest that this can be achieved, under certain conditions and assumptions, by accurately reproducing the sound field along the listener's interaural axis only. The interaural axis is that which the listener's ears lie upon. This differs to other sound field reproduction methods, such as HOA, that aim to reproduce the sound field accurately over a region of space, not just a single axis [2], [3]. The HOS approach is advantageous as it leads to less stringent requirements on the number of audio channels, loudspeakers and the loudspeaker positions compared to HOA. The approach is preferable to alternatives such as Crosstalk Cancellation (CTC) [13]–[16] which consider the sound field at the two ear positions only, as these techniques lead to more complicated frequency-dependent loudspeaker filters as opposed to simple panning gains. Furthermore, CTC makes explicit assumptions about the listener's Head-Related Transfer Function (HRTF) which are not required when using sound field reproduction approaches such as HOS.

The analysis is restricted to 2D by considering the horizontal plane only, using a coordinate system defined by radial distance, r and azimuthal angle, θ. The 2D scenario leads to symmetry which greatly simplifies the 3D case, which will be discussed after. Consider a listener with their head centred at the origin as in Fig. 1. Let $\hat{\mathbf{n}}$ be the unitary vector pointing from the head centre, $\mathbf{r}_c$, to the left ear, $\mathbf{r}_l$, thus defining the interaural axis. Assuming the listener's ears are diametrically opposed across the head and that the head radius is given by a, the two ear positions are $\mathbf{r}_{l,r} = \pm a\hat{\mathbf{n}}$. While the context is considering reproduction across the listener's interaural axis, for now in the mathematics no complex HRTF is included, a common assumption when deriving spatial audio panning laws/sound field reproduction techniques. The acoustical effects of the HRTF will dictate the performance of the technique, however this analysis will be considered in future work. It may also be noted that a valid low frequency approximation of the plane wave rigid sphere HRTF model is a shadowless head model with enlarged head radius and therefore may be modelled as two points in free field [17], which suggests at low frequencies for plane wave sources the HRTF has minimal effect. The consequence for now is the head orientation is purely used to define the position of the expansion/reproduction line and free field conditions are assumed. The head orientation is fixed such that $\hat{\mathbf{n}} = \hat{\mathbf{y}}$.

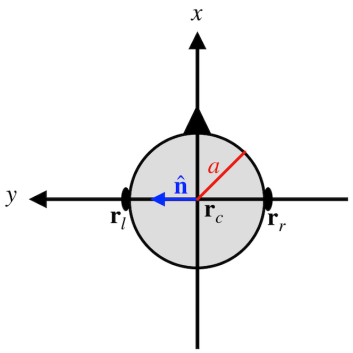

Fig. 1: Example head orientation with the interaural axis aligned along the y axis.

The incident sound source is assumed to act as a plane wave. Plane wave sources are a common assumption in the literature, and form the basis for deriving the stereo sine and stereo tangent law as well as HOA mode matching [4], [18]. The sound field due to a plane wave incident with wavevector $\mathbf{k}_i = k[\cos(\theta_i), \sin(\theta_i)]^T$, wavenumber k and measured at $\mathbf{r} = [x, y]^T$ is

$$p(\mathbf{r}) = e^{j\mathbf{k}_i \cdot \mathbf{r}} = e^{jk[x\cos(\theta_i) + y\sin(\theta_i)]}. \tag{2}$$

Next, the Taylor expansion will be used to expand the sound field about the centre of the listener's head, $r_c = [x_c, y_c]^T$, along the interaural axis, $\hat{\mathbf{n}} = \hat{\mathbf{y}}$. The n-th order term of the Taylor expansion is dependent on the n-th order derivative of the function, which for a plane wave source with respect to the y axis is

$$\frac{\partial^n}{\partial y^n} p(\mathbf{r}_c) = [jk\sin(\theta_i)]^n p(\mathbf{r}_c) \tag{3}$$

hence the Taylor expansion of the plane wave along the y axis with step size $(y - y_c)$, is

$$p(y) = \sum_{n=0}^{\infty} \frac{[jk(y-y_c)\sin(\theta_i)]^n}{n!} p(\mathbf{r}_c). \tag{4}$$

Finally, apply the expansion to be to the listener's two ears as per the head orientation and definition in Fig. 1, so that the step size is simply the head radius, $(y - y_c) = a$. Let the head be centred at the origin. Setting the plane wave to have unitary amplitude at the centre of the head implies $p(\mathbf{r}_c) = 1$.

Therefore the pressure at the listener's ear positions, $\mathbf{r}_{l,r} = \pm a\hat{\mathbf{y}}$, is given by

$$p(\pm a) = \sum_{n=0}^{\infty} \frac{[jk(\pm a)\sin(\theta_i)]^n}{n!}. \tag{5}$$

This formulation is the result of an expansion of a plane wave sound field along a line using the Taylor series as the expansion basis. The n-th order term is defined by the n-th order derivative of the sound field, which for the simple case of a plane wave and expansion along the y axis results in sine terms to the power of n. Interestingly, defining the expansion to be across the x axis (for a rotated head orientation such $\hat{\mathbf{n}} = \hat{\mathbf{x}}$) results in a similar representation except considering cosine terms to the power of n. Both approaches are equally valid, and can be transformed between by applying a rotation of the reference system. The cosine formulation is given by

$$p(\pm a) = \sum_{n=0}^{\infty} \frac{[jk(\pm a)\cos(\theta_i)]^n}{n!}. \tag{6}$$

The spatial/frequency quantity is ka, relating to the expansion step a from the origin along the line. For the context of reproduction of binaural signals, the derivation has considered the geometry of a listener's head including a head radius a and interaural axis $\hat{\mathbf{n}}$. However, the approach remains generalised to the expansion across a line utilising any finite step size or direction from the expansion centre. That is we are considering the sound field across a line, motivated by the interaural axis. To accurately represent the sound field to a higher value of ka, higher order terms of the expansion are required. On truncation of the series this fixes the ka value to which accurate representation can be achieved. Here the sound field may be considered as a line in frequency (k) or spatially (a) with the ka value setting a bound on accurate reproduction of the sound field in either domain.

The expansion terms are not strictly modes as they do not necessarily form an orthogonal basis over the expansion space, unlike for example similar work with spherical harmonics in HOA [3]. Thus from the presented Taylor expansion of a plane wave, the HOS *order* matching (as opposed to 'mode matching') equations will now be derived.

III. Higher-Order Stereophony Order Matching

A. Target and Reproduced sound fields

The target sound field is that which the loudspeaker array aims to reproduce, leading to the definition of a specific set of loudspeaker gains. The target sound field, $p_T(ka)$, is simply a plane wave as defined in (5):

$$p_T(ka) = \sum_{n=0}^{\infty} \frac{[jka\sin(\theta_T)]^n}{n!}. \tag{7}$$

A target of a single plane wave is considered for the derivation. The solution for a sound field consisting of a summation of plane waves comprises of a linear superposition of the individual plane wave contributions. Furthermore, most sound fields can be represented as a summation of plane waves, considering the plane wave density representation [19]. This reasoning follows that of HOA mode matching derivations.

The reproduced sound field is that due to the summed contributions of each individual loudspeaker in the reproduction array. Consider an array of L loudspeakers radially equidistant to the origin, which all act as plane waves. If the loudspeakers are not equidistant, a delay may be applied to them such that they are acoustically equidistant. The ℓ-th loudspeaker is situated at an angle θ_ℓ and driven by a gain g_ℓ. Considering (5), the reproduced sound field along the y axis is given by

$$p_R(ka) = \sum_{\ell=1}^{L} g_\ell \sum_{n=0}^{\infty} \frac{[jka\sin(\theta_\ell)]^n}{n!}. \tag{8}$$

B. Loudspeaker Gains Definition

The goal is to find the loudspeaker gains which minimise $||p_T(ka) - p_R(ka)||_2^2$. For exact reproduction (no order truncation) this leads to the condition $p_T(ka) = p_R(ka)$ which requires that the number of loudspeakers must be infinite. Later, the effects of truncation will be considered.

$$\sum_{n=0}^{\infty} \frac{[jka\sin(\theta_T)]^n}{n!} = \sum_{\ell=1}^{L} g_\ell \sum_{n=0}^{\infty} \frac{[jka\sin(\theta_\ell)]^n}{n!}. \tag{9}$$

Apply the order matching principle, such that the terms of the two expansions are matched for each order n. Traditionally an orthogonality condition is applied to lead to this condition. However, directly equating the n-th order terms will still result in the correct overall summation even though it may not be the only possible solution. The order matching requirement is

$$\frac{[jka\sin(\theta_T)]^n}{n!} = \sum_{\ell=1}^{L} g_\ell \frac{[jka\sin(\theta_\ell)]^n}{n!} \quad \forall\ n\ \in\ \mathbb{N}_0. \tag{10}$$

This reveals that the ka dependence of the n-th term is given by $(ka)^n$. Thus for small ka, only low order terms are required. Increasing the value of ka leads to higher order terms becoming significant. Removing all remaining common terms gives the HOS order matching equation

$$\sin^n(\theta_T) = \sum_{\ell=1}^{L} g_\ell \sin^n(\theta_\ell) \quad \forall\ n\ \in\ \mathbb{N}_0. \tag{11}$$

This means order matching with respect to powers of $\sin(x)$ leads to an accurate reproduction of the sound field along the y axis. Furthermore, truncating the expansion to a finite order N, termed N-th order stereo, only requires $L \geq N+1$ loudspeakers, less than the HOA approach. Importantly, as there is no frequency dependence in the order matching equations, the loudspeaker gains are real-valued and define simple amplitude panning laws.

To formulate the set of linear equations to find the loudspeaker gains, assume truncation to the N-th order. Let $\mathbf{p}_T$ be a length $(N+1)$ vector of target signals, Ψ be an $(N+1)\times L$ plant matrix and $\mathbf{g}$ be a length L vector of loudspeaker

gains. To define the loudspeaker gains an inverse problem is formulated:

$$\begin{aligned}
&\mathbf{p}_T = \Psi \mathbf{g} \implies \mathbf{g} = \Psi^{\dagger}\mathbf{p}_T \\
&\mathbf{p}_T = \begin{bmatrix} 1 & \sin(\theta_T) & \sin^2(\theta_T) & \dots & \sin^N(\theta_T) \end{bmatrix}^T \\
&\Psi = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ \sin(\theta_1) & \sin(\theta_2) & \sin(\theta_3) & \dots & \sin(\theta_L) \\ \sin^2(\theta_1) & \sin^2(\theta_2) & \sin^2(\theta_3) & \dots & \sin^2(\theta_L) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sin^N(\theta_1) & \sin^N(\theta_2) & \sin^N(\theta_3) & \dots & \sin^N(\theta_L) \end{bmatrix} \\
&\mathbf{g} = \begin{bmatrix} g_1 & g_2 & g_3 & \dots & g_L \end{bmatrix}^T.
\end{aligned} \tag{12}$$

The plant matrix formulates the contribution of each loudspeaker to each order, which is dictated by the angular position of the loudspeakers. The superscript $(\cdot)^{\dagger}$ indicates the Moore-Penrose pseudoinverse, a common approach for solving similar sets of linear equations in spatial audio reproduction [3], [4]. When $(N+1) \geq L$ the problem is overdetermined, an exact solution cannot be found and the pseudoinverse gives the least-squares solution that minimises the error between the target and reproduced sound field. When $(N+1) \leq L$ an infinite number of exact solutions exist, and the pseudoinverse chooses the minimum norm solution with respect to the L^2 norm.

So far, only the sine representation from expansion over the y axis has been considered. As explained in Section II-B considering reproduction across the x axis leads to a similar style solution except using a cosine formulation for $\mathbf{p}_T$ and Ψ, where the set of resulting loudspeaker gains now leads to reproduction across the x axis, not the y axis.

C. The Instability Condition

The contribution of each loudspeaker is governed by the sine or cosine of its angular position. Consider an N-th order system utilising the minimum required number of $N+1$ loudspeakers. For a pair of loudspeakers i and j situated at angles $\theta_i,\ \theta_j$ respectively, the scenario when $\sin(\theta_i) = \sin(\theta_j)$ results in both loudspeakers contributing to the reproduction axis identically and thus the system views both as 'identical' loudspeakers. That is, only N degrees of freedom are available and an exact solution can no longer be achieved. This issue arises from the cone of confusion, as the sound field recreated by the loudspeakers across a single axis only is considered. This means for a given loudspeaker at θ_i then $\theta_j \neq \pi - \theta_i$.

In the limit where $\theta_j \to \pi - \theta_i$, the plant matrix becomes ill-conditioned leading to large loudspeaker gain definitions. To overcome the ill-conditioning, an additional loudspeaker can be added at a more appropriate angular position. Alternatively, a practical method to combat the large loudspeaker gains whilst retaining the use of only $N+1$ loudspeakers is to employ Tikhonov regularisation when inverting the plant matrix [20]. This approach seeks the solution that minimises both the error between the reproduced and the target signals as well as the energy of the loudspeaker gains, weighted by a regularisation parameter. Practically this will apply a limit to the loudspeaker gains however at the cost of introducing further error into the solution. However, if the problem tends to the overdetermined scenario the exact solution already cannot be found. Thus this approach stabilises the loudspeaker gains without adding more loudspeakers but at the cost of allowing errors in the solution.

IV. Example Higher-Order Stereophony Systems

A. First Order Stereo

HOS loudspeaker gains will now be derived for a classic stereo loudspeaker setup, revealing the link of the technique to existing stereo systems. This motivates the naming as *Higher-Order Stereophony*, where it is the generalisation of the stereo theory. Consider HOS performed to just the first order. The minimum number of loudspeakers required is $L = N+1 = 2$. Let the two loudspeakers be positioned as a standard stereo pair at $\pm\theta$, with the aim to reproduce a virtual source positioned at θ_T. In this case the target pressure vector, plant matrix and loudspeaker gains are

$$\begin{aligned}
\mathbf{P}_T = \begin{bmatrix} 1 \\ \sin(\theta_T) \end{bmatrix}, \quad \Psi = \begin{bmatrix} 1 & 1 \\ \sin(\theta) & \sin(-\theta) \end{bmatrix} \\
\mathbf{g} = \frac{1}{2} \begin{bmatrix} 1 + \frac{\sin(\theta_T)}{\sin(\theta)} \\ 1 - \frac{\sin(\theta_T)}{\sin(\theta)} \end{bmatrix}.
\end{aligned} \tag{13}$$

This is traditional stereo sine law as defined in [1], [18]. Hence by using the Taylor expansion, the classic stereo sine law has been derived and is a first order Taylor approximation of reproducing the actual plane wave target sound field across a line, with the assumption the reproduction line is that of the interaural axis. The stereo sine law is defined as a low frequency approach, HOS therefore both generalises and expands the stereo theory to any given order, for any given loudspeaker array and reproduction across any frequency or spatial range (as restricted by the loudspeaker array and truncation order).

B. Second Order Stereo

With the link between classic stereo and HOS established, it is interesting to now consider a second order system. A logical step would be to consider the loudspeaker gains for a standard LCR (left-centre-right) loudspeaker setup [21]. The LR loudspeakers are a standard symmetric stereo pair, whilst the C loudspeaker is centred in front of the listener. Thus $\theta_1 = \theta_L = \theta$, $\theta_2 = \theta_C = 0$ and $\theta_3 = -\theta_L = -\theta$. This is the frontal half of a standard surround sound system (for example a 5.1, 7.1 or 5.1.2 system). For this setup with $N = 2$ the target pressure vector, plant matrix and loudspeaker gains are

$$\begin{aligned}
\mathbf{P}_T = \begin{bmatrix} 1 \\ \sin(\theta_T) \\ \sin^2(\theta_T) \end{bmatrix}, \quad \Psi = \begin{bmatrix} 1 & 1 & 1 \\ \sin(\theta) & 0 & \sin(-\theta) \\ \sin^2(\theta) & 0 & \sin^2(-\theta) \end{bmatrix} \\
\mathbf{g} = \frac{1}{2} \begin{bmatrix} \frac{\sin(\theta_T)}{\sin(\theta)} + \left(\frac{\sin(\theta_T)}{\sin(\theta)}\right)^2 \\ 2 - 2\left(\frac{\sin(\theta_T)}{\sin(\theta)}\right)^2 \\ -\frac{\sin(\theta_T)}{\sin(\theta)} + \left(\frac{\sin(\theta_T)}{\sin(\theta)}\right)^2 \end{bmatrix}.
\end{aligned} \tag{14}$$

This scenario is interesting due to how each loudspeaker contributes to each order of the reproduction. For this specific setup, the centre loudspeaker fully controls the zeroth order. The LR pair fully recreate the first order contributions, those given by the sine terms to the power of 1 where each loudspeaker of the pair has equal magnitude but opposite phase. Finally, the second order terms, those that are sine squared, are controlled by all three loudspeakers, however the LR pair works at equal magnitude in phase whilst the C loudspeaker requires a magnitude that equals the sum of the LR contributions however working in opposite phase.

This second order system is demonstrated because it is a readily available loudspeaker arrangement, used throughout the audio industry. Thus the HOS technique could be easily implemented without any new major loudspeaker arrangements needing to be adopted. This second order system, through reproducing one more order of the Taylor expansion, will recreate the sound field within an error bound along the reproduction line to a higher ka value, thus expanding traditional stereo to a higher frequency limit.

V. Relation To Higher-Order Ambisonics

A. Transformations Between sound field Representations

Upon inspection HOS is similar in nature to HOA. They are both sound field reproduction methods, HOA in 2D and 3D reproduces the correct sound field within a circle or sphere respectively, whilst HOS is correct reproduction across a line. Both techniques utilise a mathematical sound field representation to a given order, then reproduction of said expansion by matching order terms using a loudspeaker array. Increasing the truncation order of the expansion increases its validity with respect to both frequency and distance from the expansion centre. All techniques are derived using similar assumptions (primarily plane wave virtual sources and loudspeakers), and define loudspeaker panning functions. Finally, to a first order approximation HOS has been shown to be a subset of HOA, and the sound field representations are intrinsically linked [4]–[6], [8], [9]. A formal mapping between HOS and HOA will thus now be derived.

Consider the system equation from (12), $\mathbf{p}_T = \Psi\mathbf{g}$. This equation holds regardless of the expansion used to express the sound field, as long as the same representation is used to define all entries for $\mathbf{p}_T, \Psi$ and $\mathbf{g}$. Truncation to an order N is assumed. Let the superscript $(\cdot)'$ indicate truncation to the same order N but using a different sound field representation, so that $\mathbf{p}'_T = \Psi'\mathbf{g}'$. Assume an order-limited mapping between the two sound field expansions; that is under truncation to order N the mapping exists for all terms, whilst the set of basis functions used for the two representations both span the same space. The target pressures and plant matrices are related by

$$\mathbf{p}'_T = \mathbf{A}\mathbf{p}_T, \quad \Psi' = \mathbf{A}\Psi \tag{15}$$

where $\mathbf{A}$ is a matrix that expresses the transformation between the two representations. For the underdetermined case when $L \geq (N+1)$, the pseudoinverse of Ψ is used to define the gains which are a minimum norm solution. Therefore

$$\begin{aligned} \mathbf{g} &= \Psi^\dagger \mathbf{p}_T \\ \mathbf{g}' &= \Psi'^\dagger \mathbf{p}'_T \\ &= (\mathbf{A}\Psi)^\dagger \mathbf{A}\mathbf{p}_T \\ &= \Psi^\dagger \mathbf{A}^{-1}\mathbf{A}\mathbf{p}_T \\ &= \mathbf{g}. \end{aligned} \tag{16}$$

Here the identity $(\mathbf{AB})^\dagger = \mathbf{B}^\dagger\mathbf{A}^\dagger$ has been used, as well as noting that $\mathbf{A}$ is a square matrix and therefore the psueodinverse equals the standard matrix inverse, leading to $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$. The above holds for the underdetermined case as the loudspeaker gains are a minimum norm solution, which means the solution $\mathbf{g}$ to $\mathbf{p}_T = \Psi\mathbf{g}$ has zero projection onto the null-space of Ψ. In the overdetermined scenario the solution may have non-trivial elements that map to the null-space of Ψ, in this case

$$\mathbf{g} - \tilde{\mathbf{g}} = \Psi^\dagger \mathbf{p}_T, \quad \mathbf{g}' - \tilde{\mathbf{g}}' = \Psi'^\dagger \mathbf{p}'_T \tag{17}$$

with $\tilde{\mathbf{g}}$ the component of the solution that lies on the null space of Ψ. The impact of this is that for the overdetermined case the mapping can not be said to hold. Note for the underdetermined scenario the definition of a minimum norm solution is that $\tilde{\mathbf{g}} = \tilde{\mathbf{g}}' = \mathbf{0}$ which removes the issue. Hence for the underdetermined case only both representations will give identical loudspeaker gains, if and only if there is a full mapping between the terms of each expansion type. This would not hold if one term of the first expansion can not be expressed as a linear combination of the terms of the second expansion (the first expansion has a term mapped to the null space of the second expansion), then the representations are not equivalent and both will lead to differing loudspeaker gains. This result is significant as it shows that two sound field representations can be considered equivalent in the mode (or order) matching sense, and can both give identical loudspeaker gain definitions if using the minimum norm solution. As such, the goal is to determine whether such a mapping exists between any given sound field representations.

B. 2D Ambisonics To Higher Order Stereo Decoder

To consider the mapping between 2D HOA and HOS the Chebyshev polynomials are utilised. The 2D HOA sound field representation is a Fourier series, with the sound field $p(kr, \hat{\mathbf{r}})$ and $\hat{\mathbf{r}}$ dependent on the azimuthal angle θ [22]

$$p(kr,\theta) = \frac{a_0(kr)}{2} + \sum_{n=1}^{\infty} a_n(kr)\cos(n\theta) + \sum_{n=1}^{\infty} b_n(kr)\sin(n\theta). \tag{18}$$

Here $a_0(kr)$, $a_n(kr)$ and $b_n(kr)$ are coefficients found utilising the orthogonality relationships for $\cos(n\theta)$ and $\sin(n\theta)$, which form an orthogonal basis over the unit 1-sphere, $\mathcal{S}^1$ (a unit circle) [12]. The sound field across the x or y axis may be formulated by setting $\theta = 0, \pi$ or $\theta = \pi/2, 3\pi/2$ respectively. The HOS representation may be expressed using

either $\cos^n(\theta)$ or $\sin^n(\theta)$ terms each corresponding to correct reproduction across an orthogonal axis (x and y respectively). Intuitively one might expect the two sets of $\cos(n\theta)$ and $\sin(n\theta)$ terms to span the x, y axis, respectively, based on the HOS results. However, this is not necessarily the case as will now be discussed.

The goal is to find a mapping between the 2D HOA and the HOS representations when considering the sound field across only the x or the y axis in turn. For this, the Chebyshev polynomials will be used. Notably, the Chebyshev polynomials were used in a similar manner in [23] when mapping sound field derivatives measured using a differential microphone array to 2D HOA. The Chebyshev polynomials of the first kind, $T_n(x)$, expresses $\cos(n\theta)$ as a polynomial up to order n in terms of $\cos(\theta)$ [12]. These polynomials provide exactly the mapping which is required to transform between the two sound field expansions when considering the x axis expansion only, from the 2D HOA B-format representation to the equivalent HOS representation. This shows that a subset of the 2D HOA representation ($\cos(n\theta)$ terms) spans the equivalent space as the HOS representation ($\cos^n(\theta)$ terms)

As introduced in the previous section, the 2D mapping matrix $\mathbf{A}^{2D}$ is size $(N+1)\times(N+1)$ and is populated using the Chebyshev polynomial coefficients to give the mapping between HOA to HOS coefficients, having first discarded the sine terms in the HOA representation. Furthermore, expressing a plant matrix and target pressure vector in terms of $\cos(n\theta)$ or $\cos^n(\theta)$ up to the same truncation order N, using the pseudoinverse and assuming the problem is underdetermined will give identical gain definitions. This reinforces the concept that HOS ensures accurate reproduction across a single axis, and generalises previous work linking stereo and Ambisonics from first order to any given order [8], [9].

The entries of the inverse transform (HOS to HOA) given by $(\mathbf{A}^{2D})^{-1}$ are explicitly

$$A^{2D,-1}_{n',n} = t_{n',n} \text{where } T_{n'}(\cos\theta) = \sum_{n=0}^{n'} t_{n',n}\cos^n\theta \tag{19}$$

with $t_{n',n}$ the n-th coefficient of $T_{n'}$ whilst due to the nature of the Chebyshev generating functions explicitly stating the HOA to HOS transform entries (for $\mathbf{A}^{2D}$) is non trivial. An example of the order $N=2$ mapping matrix for both the forward and inverse transform is

$$\mathbf{A}^{2D} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}, \quad (\mathbf{A}^{2D})^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 2 \end{pmatrix}. \tag{20}$$

These matrices will be lower triangular, which is a consequence of the mapping being order-limited (the n-th term of one representation is given by terms to order n of the second representation, observable in (19)).

Interestingly the mapping does not exist if the HOS system is expressed in terms of $\sin^n(\theta)$. This is significant, as the HOS $\sin^n(\theta)$ representation is sufficient to represent the sound field on the y axis, however the set of $\sin(n\theta)$ from the 2D HOA representation is not, unlike with the corresponding $\cos^n(\theta)$ and $\cos(n\theta)$ scenario. Therefore, a decoder does not exist to map from this subset of 2D HOA to the HOS sine representation, even though in the eyes of HOS the sine representation is as valid as the cosine form. This may be intuitively observed from (18). Consider evaluation across the x axis, such that $\theta = 0, \pi$, then

$$p(kr, \theta = 0, \pi) = \frac{a_0(kr)}{2} + \sum_{n=1}^{\infty} a_n(kr)(\pm 1)^n \tag{21}$$

and only the coefficients corresponding to $\cos(n\theta)$ functions are required to represent the sound field across the x axis. Thus this set of functions span the same space as the HOS cosine representation with the transformation between the two given by Chebyshev polynomials of the first kind. However, when considering the y axis such that $\theta = \pi/2, 3\pi/2$ then

$$p\left(kr, \theta = \frac{\pi}{2}, \frac{3\pi}{2}\right) = \frac{a_0(kr)}{2} + \sum_{n=2,\ \text{n even}}^{\infty} a_n(kr)(-1)^{\frac{n}{2}} + \sum_{n=1,\ \text{n odd}}^{\infty} b_n(kr)(\mp 1)^{\frac{n-1}{2}}. \tag{22}$$

Thus using the 2D HOA representation the sound field across the y axis requires the $\cos(n\theta)$ coefficients and the $\sin(n\theta)$ coefficients when n is odd and even respectively. No clear mapping then exists between the HOS sine representation which covers the y axis representation of the sound field.

This means in practice, a decoder from 2D HOA to HOS first requires rotation of the 2D HOA sound field to ensure the x axis aligns with the listener's interaural axis (to pick out the $\cos(n\theta)$ terms), followed by multiplication of the HOA basis weighting coefficients by the decoding matrix as defined by the Chebyshev polynomials of the first kind. The impact of the existence of this decoder is substantial. It means that all 2D HOA content is able to be rendered over a HOS system. Whilst this does result in discarding some information about the sound field, the benefit is that 2D HOA requires a minimum of $2N+1$ loudspeakers, whilst HOS requires just $N+1$. Discarding these sound field coefficients corresponds to enforcing a cone of confusion about the y axis, as will be explored in Section VI.

C. 3D Ambsionics To Higher Order Stereo Decoder

A similar decoder from 3D HOA to HOS may also be derived. However, first the spherical harmonic expansion of a plane wave sound field must be manipulated to reveal the relationship between the two techniques. This will involve deriving the representation of a plane wave across the z axis only by utilising the spherical harmonic expansion, then performing mode matching using a subset of spherical harmonics to define the final decoder.

Consider the pressure due to a plane wave incident with wavevector and wavenumber $\mathbf{k}_i = k\hat{\mathbf{k}}_i$ at a point $\mathbf{r} = r\hat{\mathbf{r}}$ is given by $p(kr, \hat{\mathbf{r}}) = e^{j\mathbf{k}_i\cdot\mathbf{r}}$. Note a 3D coordinate system is now used and the unit vector $\hat{\mathbf{r}}$ denotes the angular dependence through the azimuth and colatitude angles ϕ_i and θ_i. The

3D Jacobi-Anger expansion expresses the plane wave as a summation of spherical harmonics [22]

$$p(kr, \hat{\mathbf{r}}) = e^{j\mathbf{k}_i \cdot \mathbf{r}} = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} 4\pi j^n j_n(kr) Y_n^m(\hat{\mathbf{k}}_i) Y_n^m(\hat{\mathbf{r}})^*. \tag{23}$$

with j_n the n-th spherical Bessel function and the direction of arrival of the plane wave being given by $\hat{\mathbf{k}}_i$. The spherical harmonics are a set of functions that form an orthonormal basis over the unit 2-sphere, $\mathcal{S}^2$ (a unit sphere). Hence, any square-integrable well-behaved function on a sphere may be expressed as a weighted linear summation of spherical harmonics. The spherical harmonic Y_n^m, of order n and degree m, may be defined in complex form as [24]

$$Y_n^m(\theta, \phi) = \sqrt{\frac{(2n+1)}{4\pi} \frac{(n-m)!}{(n+m)!}} P_n^m(\cos\theta) e^{jm\phi} \tag{24}$$

where P_n^m is the associated Legendre polynomial.

Using the spherical harmonic addition theorem the Jacobi-Anger expansion may be expressed purely in terms of Legendre polynomials [22]

$$p(kr, \hat{\mathbf{r}}) = e^{j\mathbf{k}_i \cdot \mathbf{r}} = \sum_{n=0}^{\infty} j^n (2n+1) j_n(kr) P_n(\hat{\mathbf{k}}_i \cdot \hat{\mathbf{r}}) \tag{25}$$

noting that $\hat{\mathbf{k}}_i \cdot \hat{\mathbf{r}} = \cos(\Theta)$, where Θ is the angle between $\hat{\mathbf{k}}_i$ and $\hat{\mathbf{r}}$, and P_n is the n-th order Legendre polynomial.

Next, a slight change to the coordinate system is required. In the literature it is common to align the wavevector of the incident plane wave with the z axis such that $\mathbf{k}_i = k\hat{\mathbf{z}}$, in which case $\mathbf{k}_i \cdot \mathbf{r} = kr\cos(\theta)$ with θ the colatitude. Instead, align $\hat{\mathbf{r}}$ with the z axis such that $\mathbf{r} = r\hat{\mathbf{z}}$ and the dot product $\mathbf{k}_i \cdot \mathbf{r} = kr\cos(\theta_i)$. This fixes the coordinates the sound field can be evaluated at to positions with $\theta = 0, \pi$ which with $r \in [0, \infty)$ spans the whole z axis.

Denote the positive and negative halves of the z axis with subscripts $+, -$. For the evaluation positions on the positive and negative z axis respectively, $\hat{\mathbf{k}}_i \cdot \hat{\mathbf{r}}_+ = \cos(\theta_i)$ and $\hat{\mathbf{k}}_i \cdot \hat{\mathbf{r}}_- = \cos(\pi - \theta_i) = -\cos(\theta_i)$. Utilising the parity of the Legendre polynomials [22] the plane wave may be expressed as

$$p_{+,-}(kr, \hat{\mathbf{z}}) = \sum_{n=0}^{\infty} (\pm 1)^n j^n (2n+1) j_n(kr) P_n(\cos\theta_i). \tag{26}$$

As the orthogonality of the spherical Bessel functions is later required, their argument must be extended to cover the region $(-\infty, \infty)$. Thus define a change in coordinate system

$$\begin{aligned} r \in [0, \infty) &\implies r' \in (-\infty, \infty) \\ \theta = 0, \pi &\implies \theta = 0 \end{aligned} \tag{27}$$

with the sound field represented as

$$p(kr') = \begin{cases} \sum_{n=0}^{\infty} j^n (2n+1) j_n(kr) P_n(\cos\theta_i) & \text{if } kr' \geq 0 \\ \sum_{n=0}^{\infty} (-1)^n j^n (2n+1) j_n(kr) P_n(\cos\theta_i) & \text{if } kr' < 0 \end{cases}. \tag{28}$$

Note that the expression for $p_-(kr)$ is the same as $p_+(kr)$ except for the additional term $(-1)^n$. In the new coordinate system using r', this factor of $(-1)^n$ may be absorbed back into the spherical Bessel functions of argument kr' using the property $j_n(-x) = (-1)^n j_n(x)$ [25]. Finally,

$$p(kr') = \sum_{n=0}^{\infty} j^n (2n+1) j_n(kr') P_n(\cos\theta_i). \tag{29}$$

Crucially, in performing this rotation and fixing the evaluation of the equation to along the z axis, only the zonal spherical harmonics are required. These are the spherical harmonics with $m = 0$ which have no dependence on the azimuthal angle ϕ. It may be observed that these spherical harmonics form a basis for all axisymmetric functions on a sphere which have no azimuthal dependence. Thus from the full set of spherical harmonics only the following are utilised:

$$Y_n^0(\theta, \phi) = \sqrt{\frac{(2n+1)}{4\pi}} P_n(\cos\theta). \tag{30}$$

Hence, instead of considering $(N+1)^2$ spherical harmonics to the N-th order, now only $(N+1)$ play a role. Thus the 3D HOA sound field representation has been manipulated using a rotation to consider a subset of spherical harmonics that represent the sound field along the z axis only.

1) Target and Reproduced sound fields: A set of mode matching equations will now be defined but instead using the expansion in (29). The target sound field, $p_T(kr')$ is that of a plane wave and given by

$$p_T(kr') = \sum_{n=0}^{\infty} j^n (2n+1) j_n(kr') P_n(\cos\theta_T). \tag{31}$$

For a reproduction array of L equidistant loudspeakers, that act as plane waves with the ℓ-th loudspeaker making an angle θ_ℓ with the z axis and being driven by a gain g_ℓ, the reproduced sound field, $p_R(kr')$, is

$$p_R(kr') = \sum_{\ell=1}^{L} g_\ell \sum_{n=0}^{\infty} j^n (2n+1) j_n(kr') P_n(\cos\theta_\ell). \tag{32}$$

2) Loudspeaker Gains Definition: As before, the aim is to find the loudspeaker gains which lead to accurate reproduction of the target sound field. Begin by equating $p_T(kr') = p_R(kr')$, that is equating (31) and (32). Unlike with normal HOA mode matching, here is no colatitude or azimuthal angle to integrate over and thus no corresponding angular dependent function for which an orthogonality condition can be exploited. Instead, make use of the orthogonality of the spherical Bessel functions over the region $(-\infty, \infty)$, corresponding to the fact that the region being considered is an infinite line (the z axis).

The spherical Bessel functions form an orthogonal basis over the region $[-\infty, \infty]$ [26]

$$\int_{-\infty}^{\infty} j_n(x) j_{n'}(x) dx = \frac{\pi}{(2n+1)} \delta_{nn'}. \tag{33}$$

Multiplying both sides of the equality by a dummy variable $j_{n'}(kr')$, integrating over $[-\infty, \infty]$, using the spherical Bessel function orthogonality condition then removing common terms leads to the relation

$$P_n(\cos\theta_T) = \sum_{\ell=1}^{L} g_\ell P_n(\cos\theta_\ell). \tag{34}$$

This is an interesting form of mode matching equation, dependent on mode matching the Legendre polynomials. Due to this specific rotation to align the evaluation along the z axis, this actually equates to correct reproduction along the z axis only. This bares striking resemblance to HOS.

3) Decoder Definition: Armed with the new mode matching approach in (34), all that is needed to decode from the 3D Ambisonics representation to HOS is a mapping between the two sound field representations. This is in fact very simple, as the n-th order Legendre polynomial is exactly a polynomial in terms of $\cos(\theta)$ to the n-th order by definition. Therefore, the coefficients of the Legendre polynomials fill the entries of the mapping matrix $\mathbf{A}^{3D}$ to decode from the relevant subset of the 3D Ambisonics representation (the space spanned by the spherical harmonics with $m = 0$) to the cosine HOS representation. Furthermore as proven earlier, because such a mapping exists the gain definitions from mode matching the Legendre polynomials in this manner will be exactly the same as those from the HOS approach (when using the pseudoinverse and assuming $L \geq N + 1$).

The entries of the inverse transform (HOS to HOA) given by $(\mathbf{A}^{3D})^{-1}$ are explicitly

$$A^{3D,-1}_{n',n} = l_{n',n} \text{where } P_{n'}(\cos\theta) = \sum_{n=0}^{n'} l_{n',n} \cos^n\theta \tag{35}$$

with $l_{n',n}$ the n-th coefficient of $P_{n'}$. An example of such of the decoder matrix up to order $N = 2$ is

$$\mathbf{A}^{3D} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \frac{1}{3} & 0 & \frac{2}{3} \end{pmatrix}, \quad (\mathbf{A}^{3D})^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{1}{2} & 0 & \frac{3}{2} \end{pmatrix}. \tag{36}$$

As with the 2D case, these matrices will be lower triangular. The rotation is key for the definition of the decoder. The sound field must be rotated so that the HOS expansion axis is along the z axis, or equivalently that θ is the angle from both the z and interaural axis. This is the only rotation that will pick out the set of $(N + 1)$ spherical harmonics with $m = 0$ that can represent the sound field along one axis only. This may be viewed as actually rotating the sound field so that the listener's interaural axis lies along the z axis.

Therefore, a 3D Ambisonics to cosine HOS decoder exists in a similar manner to the 2D Ambisonics decoder and is shown as a signal flow in Fig. 2. First, the Ambisonics sound field must be rotated such that the interaural axis aligns with the z axis, using standard approaches for the rotation of B-Format signals [4]. Then a subset of the 3D Ambisonic B-format signals must be multiplied by a matrix $\mathbf{A}^{3D}$ whose entries are defined by the Legendre polynomials. As with the 2D Ambisonics decoder, the 3D Ambisonics decoder means all 3D Ambisonics content can be rendered over a HOS system, using only $(N + 1)$ loudspeakers as opposed to $(N + 1)^2$.

VI. Formulation For Elevated Sources

So far the HOS approach has been defined using a 2D coordinate system. However, the existence of the 3D HOA to HOS decoder suggests the technique is also applicable for virtual source positions in 3D. In Section III-C the instability condition demonstrated how, by considering the sound field reproduced across a single line only, a loudspeaker in front or behind the listener are viewed as identical by the HOS system. This is the scenario when $\sin(\theta_i) = \sin(\theta_j)$ which is satisfied when $\theta_i = \pi - \theta_j$, and is the 2D equivalent of the cone of confusion. Whilst this creates a limitation on the reproduction loudspeaker positions, it can be taken advantage of when considering the virtual target source. That is a virtual source behind the listener can be represented as a virtual source in the frontal region, as both lead to an identical sound field across the analysis axis and therefore identical loudspeaker gains.

Now consider the 3D scenario where the virtual source is elevated. The cone of confusion about a given axis is defined as all positions which have the same angle measured from the interaural axis [27]. Therefore, in a similar manner to the 2D case, a source with elevation can be mapped to a source position in the frontal horizontal plane (with no elevation) through the cone of confusion, as both positions will lead to the same sound field across the evaluation axis. This holds when the free field and plane wave assumptions made in the HOS derivation are satisfied. In this case, consider using a 3D coordinate system with evaluation across the y axis, defined by $\phi_y = \pi/2, \theta_y = \pi/2$. Let $\hat{\mathbf{r}}_e, \hat{\mathbf{r}}_h$ be the desired elevated source position and the equivalent horizontal only position mapped through the cone of confusion respectively, which is illustrated in Fig. 3. We will now derive the equivalent horizontal only source position, with ϕ_h to be determined and $\theta_h = \pi/2$, that maps from $\hat{\mathbf{r}}_e$ to $\hat{\mathbf{r}}_h$ using the cone of confusion. Begin with the scalar product and the great circle distance that states between two positions on a circle

$$\cos(\Theta) = \cos(\theta_1)\cos(\theta_2) + \sin(\theta_1)\sin(\theta_2)\cos(\phi_1 - \phi_2). \tag{37}$$

Θ is defined as the angle between $\hat{\mathbf{r}}_e$ and $\hat{\mathbf{r}}_h$. The cone of confusion requires the angle of the two source positions from the y axis to be equal, $\cos(\Theta_h) = \cos(\Theta_e)$, this leads to

$$\phi_h = \arccos\left[\sin(\theta_e)\sin(\phi_e)\right] + \frac{\pi}{2}. \tag{38}$$

This relationship maps any elevated source position to the equivalent horizontal only position that leads to the same sound field across the y axis. Therefore, the resulting HOS

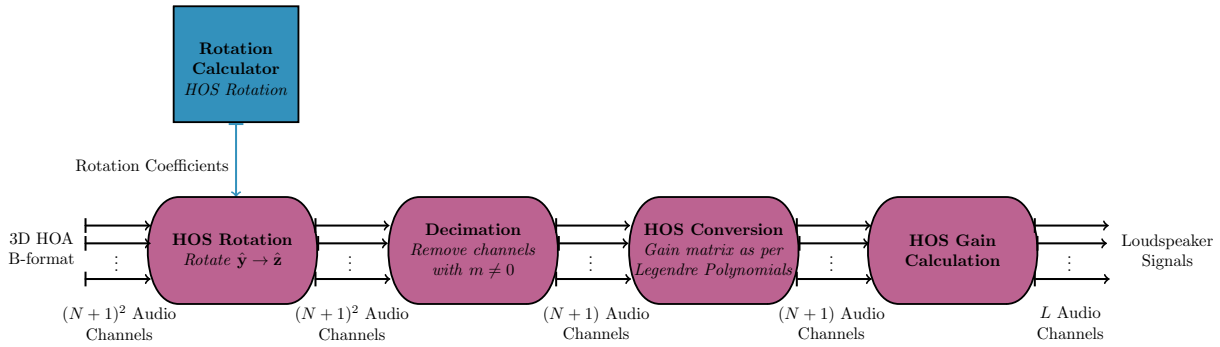

Fig. 2: Signal flow for reproducing 3D HOA B-format utilising HOS.

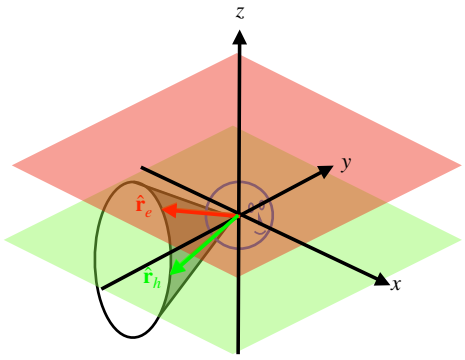

Fig. 3: Cone of confusion mapping of an elevated position ($\hat{\mathbf{r}}_e$ with θ_e, ϕ_e) to one with no elevation ($\hat{\mathbf{r}}_h$ with $\theta_h = \pi/2, \phi_h$). The planes indicate positions with equal elevation.

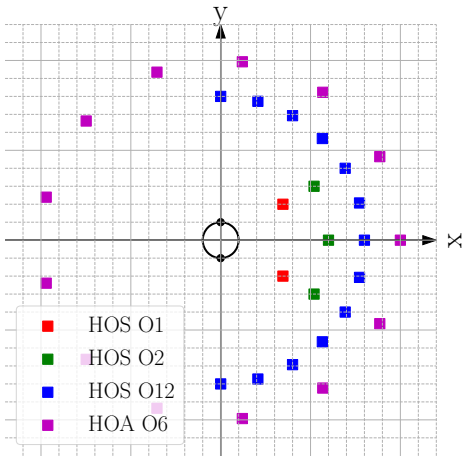

Fig. 4: Loudspeaker layouts for the three HOS and one 2D HOA system. The plot indicates angular arrangements only, as all loudspeaker were positioned at the same radial distance.

loudspeaker gains will be identical for both positions. In this sense, HOS will reproduce any elevated source position through an equivalent horizontal only position, using horizontal only loudspeakers. However, elevation specific cues such as pinna notches will not be reproduced as the assumption holds only when the cone of confusion is valid, which is true for low frequencies [27].

VII. Experimental Validation

Measurements were performed to validate the analytical results derived so far for the new proposed HOS technique. The aim of the measurements was to collect a database of transfer functions in anechoic conditions from a reference source to a microphone array, with the source at a fixed radial distance but measuring for different angular positions in the horizontal plane. From these measurements, the reproduced sound field due to any given arrangement of loudspeakers in the horizontal plane can then be simulated. Measurements were made in the horizontal plane only and therefore a 2D HOA approach was considered.

Four systems were compared and are detailed in Fig. 4 and Table I. Three HOS systems with increasing truncation orders (HOS O1, O2 and O12) were analysed to investigate if increasing the order did yield increased accuracy in the reproduced sound field at higher frequencies/larger distances from the reproduction point. HOS O12 with thirteen loudspeakers spaced equally across a semicircle in front of the listener was

Approach	Truncation Order	Number of Loudspeakers
HOS	1	2
HOS	2	3
HOS	12	13
HOA	6	13

TABLE I: Details of the systems under comparison.

then chosen to compare to a reference sixth order 2D HOA rig (HOA O6), which also requires thirteen loudspeakers but now spaced equally across a circle surrounding the listener. These systems were chosen as when using the same number of loudspeakers HOS can achieve a higher truncation order than HOA, as well as considering that the HOS approach requires loudspeakers in front of the listener only.

All measurements were performed in the large anechoic chamber at the Institute of Sound and Vibration Research (ISVR), University of Southampton, to ensure freefield conditions. The experimental apparatus was the same for all measurements except for changing the microphone array. Each microphone array was mounted in turn on a turntable. A single Genelec 8020C loudspeaker was used as a reference sound source, positioned 3 meters from the microphone array to approximate a plane wave source. Transfer functions were measured from the loudspeaker to the microphone array using

exponential sine sweeps for all horizontal angular positions to a 1 degree resolution. To remove reflections from any equipment in the anechoic chamber, an adapted approach of frequency-dependent windowing was applied as detailed in [9].

Two different microphone arrays were utilised. The first was a linear array of 15 B&K type 4189 omnidirectional microphones spaced with 0.037 m separation between each microphone. This microphone spacing results in a spatial aliasing frequency of approximately 4600 Hz. The linear array was used to sample the sound field across a line, to investigate the claim that HOS results in correct reproduction across a given axis. A calibration procedure ensured the response of all array channels were properly matched. The second microphone array was an Eigenmike EM32 fourth order HOA microphone, a 32 capsule spherical microphone array that can be used to sample the spherical harmonic coefficients of a sound field [28], [29]. This array was measured to consider the contribution of a HOS system to the spherical harmonic modes of the reproduced sound field, to verify whether HOS does accurately reproduce the $m = 0$ modes in the rotated reference coordinate system only as claimed through the 3D HOA to HOS decoder derivation. The raw output of the Eigenmike resulted in measurements from the loudspeaker to each of the microphone capsules. Thus the accompanying Eigenmike software was used to convert these measurements to B-format signals, to obtain the impulse responses from the loudspeaker to each B-format channel. This was performed up to order $N = 4$. The B-format impulse responses then underwent a rotation to align the z axis of the spherical harmonic expansion basis with the reproduction axis (the y axis).

A. Linear Array Results

The reproduced sound field at each microphone position was simulated using $\mathbf{p}_R = \Psi^{measured}\mathbf{g}$. Here, $\mathbf{p}_R^{measured}$ is the vector of pressures at the microphone positions, in this case sampling the sound field across the reproduction axis (the y axis). The target sound field, $\mathbf{p}_T^{measured}$, was defined as the real measurement of the loudspeaker at the required virtual source position. The measured transfer functions fill the entries of the plant matrix, $\Psi^{measured}$, and the loudspeaker gains, $\mathbf{g}$, are specified as per the desired order HOS or HOA approach. The normalised complex error as a function of frequency, $\epsilon(f)$, is defined as the normalised difference between the reproduced and target sound fields, such that for each microphone position

$$\epsilon(f) = \frac{|p_T^{measured}(f) - p_R^{measured}(f)|^2}{|p_T^{measured}(f)|^2}. \tag{39}$$

This error metric takes into account both magnitude and phase differences between the reproduced and target field. The error is considered using a decibel scale, where a smaller value indicates less error in which case the sound field reproduced by the system better matches that of the target.

Fig. 5 shows the complex error across the array length (x axis on the plots) and as a function of frequency for each of the reproduction systems for $\theta_T = 10°$. The red dotted lines indicate the $N = kr$ limit for each of the systems [3], under which it is expected that all systems should perform well with little error in this region. These results confirm that increasing the order of the HOS approach leads to more accurate reproduction with respect to the kr quantity, as little error is observed within the $N = kr$ limit whilst outside of it maximal error occurs. Therefore utilising a higher order system leads to more accurate reproduction at higher frequencies and across a larger distance across the y axis. In general, HOS appears to follow this $N = kr$ rule of thumb originally derived for HOA. If the number of loudspeakers is fixed the HOS approach is advantageous to the HOA technique. This is because a higher order of reproduction can be achieved leading to a larger region of validity across the reproduction axis, whilst also only requiring loudspeakers in front of the listener making for a more accessible loudspeaker array. Notably, even within the $N = kr$ bounds there is some error at high frequencies, due to spatial aliasing dictated by the microphone array spacing.

Fig. 6 shows the complex error for all the systems across the microphone array length, however now as a function of all virtual source positions. Individual frequencies are shown for $250, 1000$ and 2000 Hz. Once more it is apparent that increasing the order of the system leads to an increased area of accurate reproduction across the y axis. As all systems use the minimum number of loudspeakers required, both the HOS and HOA gain definitions activate a single loudspeaker when the virtual source is positioned at that given loudspeaker. Therefore, lines of zero error are apparent throughout the results when θ_T is at a loudspeaker position. This also reveals how the HOS technique takes advantage of the instability condition explained in Section III-C, as due to the cone of confusion the sound field due a virtual source along the analysis axis is equal for θ_T and $\theta_T' = 180° - \theta_T$. This can be viewed as a mirroring operation about the analysis axis. Therefore at loudspeaker positions the sound field is also correct for the mirrored position on the cone of confusion. For example with HOS O1 the loudspeakers are positioned at $\pm 30°$, thus correct reproduction is observed when $\theta_T = \pm 30°$ and $\pm 150°$. Finally, comparing HOS O12 and HOA O6 it is clear that as the frequency increases, the sound field is reproduced correctly over a larger distance on the evaluation axis for the HOS technique due to it working to a higher truncation order.

Finally, Fig. 7 illustrates the complex error across the whole sound field in the interior of the loudspeaker array for a single frequency and virtual source position (2000 Hz, $\theta_T = 68°$). This demonstrates that HOS, whilst reproducing the sound field accurately along the reproduction line, does not reproduce the sound field over a wider circular area, unlike with HOA. For HOA O6 some error is seen within the $N = kr$ boundaries which is most likely experimental error, however, broadly the sound field is correct across this region. For HOS O1 the accurate reproduction region is very small at this frequency, whilst for HOS O2 a slightly larger sweet spot is observed with some minimal error off the reproduction axis. This suggests some level of robustness to misalignment of the listener's interaural axis with the reproduction axis. HOS O12 exhibits a significant area of correct reproduction close to the interaural axis, thus here a significant amount of robustness to listener misalignment is expected.

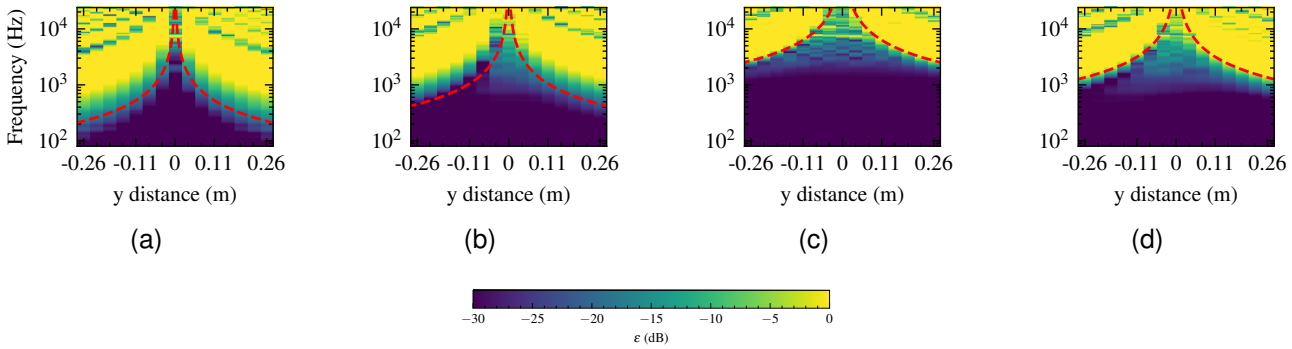

Fig. 5: Complex error of the reproduced sound field, across frequency and distance along the y axis for $\theta_T = 10°$. The red dotted lines indicate the $N = kr$ limit. Columns varying renderer type (HOS O1, HOS O2, HOS O12, HOA O6 respectively).

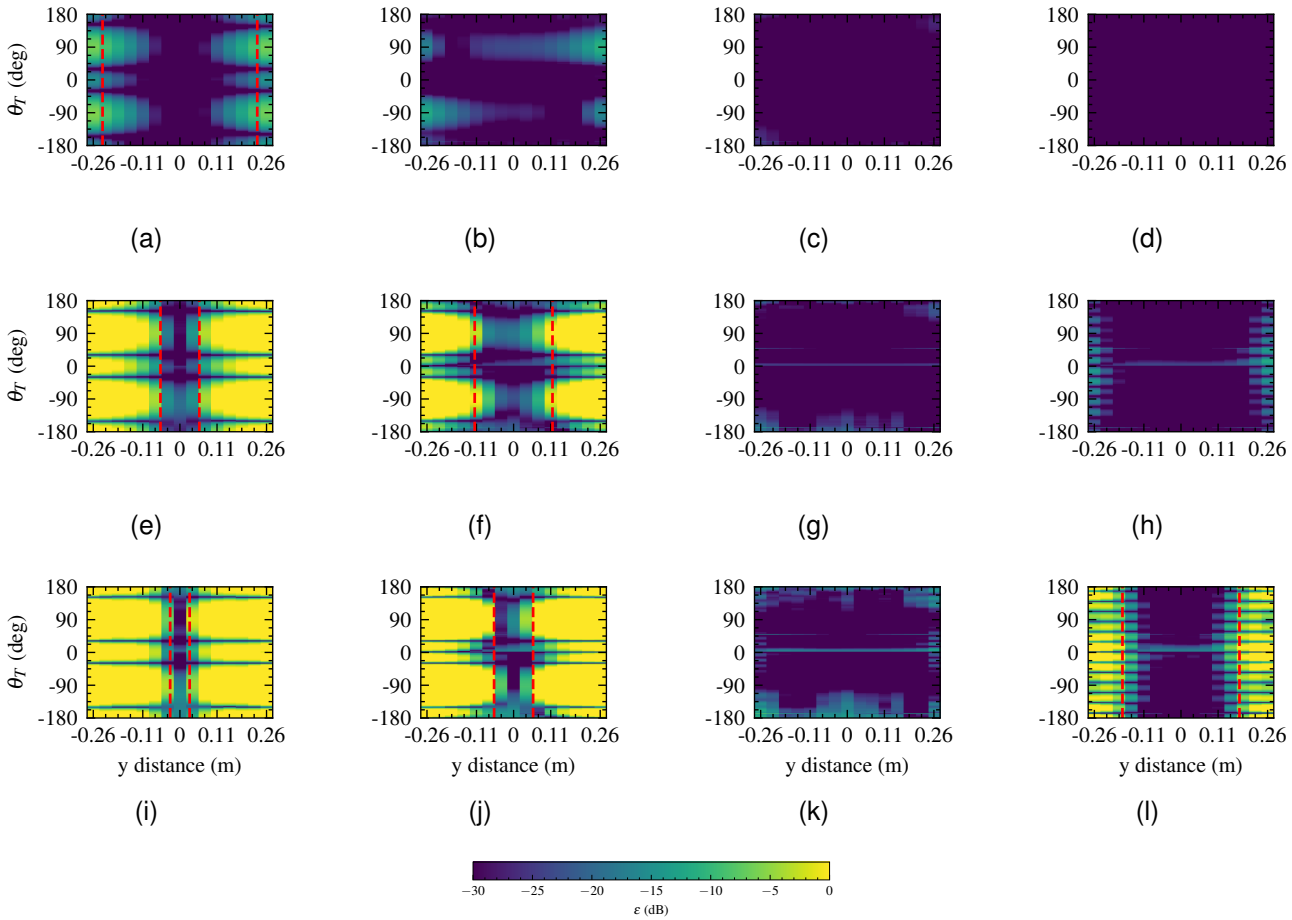

Fig. 6: Complex error for of the reproduced sound field, across all virtual source positions and distance along the y axis for fixed frequencies. The red dotted lines indicate the $N = kr$ limit. Rows varying frequency (a-d) 250 Hz, (e-f) 1000 Hz, (i-l) 2000 Hz. Columns varying renderer type (HOS O1, HOS O2, HOS O12, HOA O6 respectively).

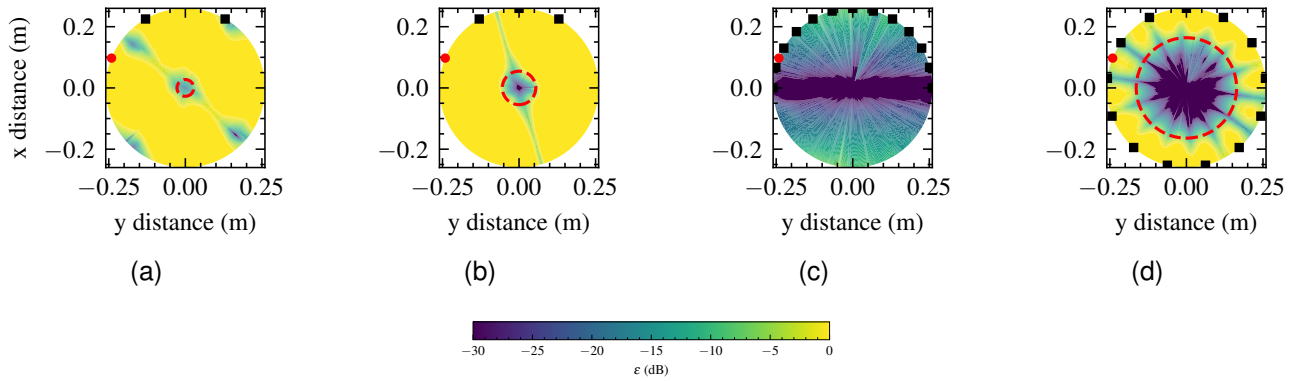

Fig. 7: Complex error of the reproduced sound field at 2000 Hz for $\theta_T = 68°$. The red dotted lines indicate the $N = kr$ limit, whilst the red circle indicates the virtual source position and the squares show the loudspeaker positions. Columns varying renderer type (HOS O1, HOS O2, HOS O12, HOA O6 respectively).

B. Eigenmike Results

A similar approach was used for the Eigenmike except the plant matrix was populated with the loudspeaker to B-format transfer functions, with the vector $\mathbf{p}_R^{measured}$ now the reproduced spherical harmonic coefficients of order $n \in [0, 4], m = 0$ only. Only the $m = 0$ spherical harmonics are considered as following the rotation to align the z axis with the reproduction axis, only the $m = 0$ subset is required to represent the sound field along this axis. However, notably the sound field reproduced using HOS will be incorrect when deviating from this line, unlike with HOA. The complex error was again calculated between the reproduced and target sound fields. The normalisation term was chosen as the measured target $n, m = 0, 0$ coefficient, $W_T^{measured}$ (an omnidirectional microphone response).

$$\epsilon(f) = \frac{|p_T^{measured}(f) - p_R^{measured}(f)|^2}{|W_T^{measured}(f)|^2}. \tag{40}$$

Fig. 8 shows the complex error for the reproduced spherical harmonic coefficients corresponding to each B-format channel, as a function of frequency and virtual source angular position. The Eigenmike introduces spatial aliasing above 6000 Hz, therefore there is considerable error in this region regardless of the technique. The results demonstrate that HOS does indeed accurately reproduce the $m = 0$ coefficients up to the truncation order. This experimentally verifies the link established between HOS and HOA through the decoder definition, and that controlling the $m = 0$ channels only corresponds to accurate sound field reproduction along a single axis. The HOS approaches all encounter issues when the virtual source is positioned behind the listener where there are no loudspeakers positioned. This suggests that whilst rear virtual sources can be achieved with HOS and just frontal loudspeakers, they may not be reproduced as robustly when compared to using HOA. However, HOA requires a fully surrounding loudspeaker array for optimal performance, unlike HOS. Interestingly, even above the truncation order HOS reproduces the spherical harmonic coefficient correctly up to approximately 800 Hz, which can be observed for HOS O1, O2 and channels $n = 3, 4$.

VIII. Conclusions

This article has introduced the theoretical foundations for a new sound field reproduction technique named Higher-Order Stereophony (HOS). HOS is founded on the Taylor expansion of the sound field due to an incident plane wave, across one axis only. This expansion represents the sound field as an infinite summation of the sound fields derivatives evaluated about an expansion point. Thus HOS leads to accurate sound field reproduction across a line only. The resulting loudspeaker gains are panning functions and the stereo sine law has been shown to be a first order HOS system. This motivates the name of the technique, where HOS generalises this classic audio reproduction approach to higher orders.

Decoders from both the 2D and 3D HOA representations to HOS have been derived. Both decoders first require a rotation to align the interaural axis across the x or z axis in 2D and 3D, respectively. The 2D HOA to HOS decoder then utilise the Chebyshev polynomials, whilst the 3D HOA to HOS decoder uses a subset of spherical harmonics with $m = 0$. Importantly, N-th order HOS requires only $(N+1)$ channels/reproduction loudspeakers, whilst 2D and 3D HOA need $(2N + 1)$ and $(N + 1)^2$, respectively. Ideally, HOA requires loudspeakers uniformly distributed over a circle or sphere. In contrast, HOS can use loudspeakers in front of the listener only, if desired.

Experimental validation was carried out using two different microphone arrays, a linear microphone array, to evaluate the reproduced sound field across a line, and a spherical microphone array, to consider the reproduced spherical harmonic coefficients of the sound field. These results have confirmed that HOS reproduces the sound field accurately across a line, following the $N = kr$ rule. HOS also correctly reproduces the $m = 0$ spherical harmonic coefficients of the sound field.

A dynamic version of HOS that is adaptive to listener movements and head rotations has been developed by the authors and will be presented in future publications. This dynamic version of HOS leads to listener-adaptive sound field reproduction and will be shown to encompass other classic stereo techniques such as the tangent law. HOS will also be analysed with the inclusion of more complex HRTFs, leading

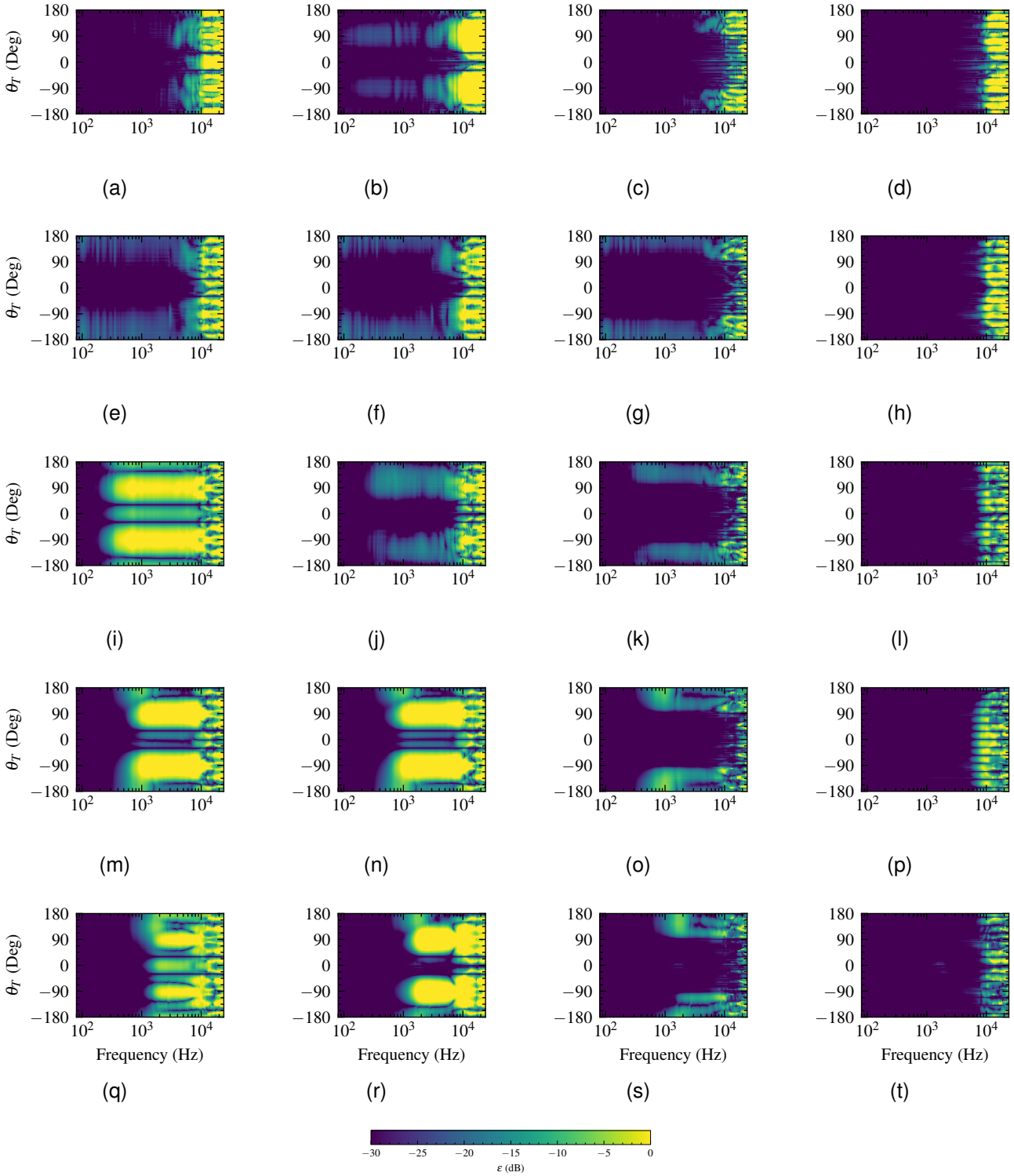

Fig. 8: Complex error of the reproduced degree $m = 0$ spherical harmonic coefficients as a function of frequency and virtual source position. Rows each show a different order spherical harmonic ($n = 0, 1, 2, 3, 4$). Columns varying renderer type (HOS O1, HOS O2, HOS O12, HOA O6 respectively).

to a new binaural rendering approach based on HOS. The observation that HOS follows the $N = kr$ rule is also to be formulated mathematically, and the inclusion of point sources for the reproduction problem investigated with the expectation this will lead to frequency-dependent loudspeaker filters as opposed to panning gains, as in HOA [30].

ACKNOWLEDGMENTS

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) through the University of Southampton's Doctoral Training Partnership under Grant 2106106. For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

REFERENCES

[1] H. A. M. Clark, G. F. Dutton, and P. B. Vanderlyn, "The 'stereosonic' recording and reproducing system. a two-channel system for domestic tape records," *Proceedings of the IEE - Part B: Radio and Electronic Engineering*, vol. 104, no. 17, pp. 417–432, 09 1957.

[2] M. A. Poletti, "Three-dimensional surround sound systems based on spherical harmonics," *J. Audio Eng. Soc*, vol. 53, no. 11, pp. 1004–1025, 2005.

[3] D. B. Ward and T. D. Abhayapala, "Reproduction of a plane-wave sound field using an array of loudspeakers," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 6, pp. 697–707, 2001.

[4] F. Zotter and M. Frank, *A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*. Springer International Publishing, 2019, vol. 19.

[5] G. Dickins and R. Kennedy, "Towards optimal soundfield representation," in *Audio Engineering Society Convention 106*, 1999.

[6] G. Dickins, "Soundfield representation, reconstruction and perception," Ph.D. dissertation, Research School of Information Sciences and Engineering, The Australian National University, 2003.

[7] D. Menzies and F. M. Fazi, "A theoretical analysis of sound localization, with application to amplitude panning," in *Audio Engineering Society Convention 138*, 05 2015.

[8] J. Hollebon and F. M. Fazi, "Generalised low frequency 3d audio reproduction over loudspeakers," in *AES 148th Convention*, 2020.

[9] ——, "Experimental study of various methods for low frequency spatial audio reproduction over loudspeakers," in *I3DA: International Conference on Immersive and 3D Audio*, 2021.

[10] D. Menzies, M. F. S. Gálvez, and F. M. Fazi, "A low-frequency panning method with compensation for head rotation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 2, pp. 304–317, 2018.

[11] D. Menzies and F. M. Fazi, "A complex panning method for near-field imaging," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 9, pp. 1539–1548, 2018.

[12] G. B. Arfken and H. J. Weber, *Mathetical Methods For Physicists*, sixth edition ed. Boston: Academic Press, 2005.

[13] P. Damaske, "Head-related two-channel stereophony with loudspeaker reproduction," *The Journal of the Acoustical Society of America*, vol. 50, no. 4B, pp. 1109–1115, 1971.

[14] M. F. Simón Gálvez and F. M. Fazi, "Loudspeaker arrays for transaural reproduction," in *The 22nd International Congress of Sound and Vibration, Florence*, 2015.

[15] J. Hollebon, F. M. Fazi, and M. F. Simón Gálvez, "A multiple listener crosstalk cancellation system using loudspeaker dependent regularization," *J. Audio Eng. Soc*, vol. 69, no. 3, pp. 191–203, 2021.

[16] M. F. Simón Gálvez, D. Menzies, and F. M. Fazi, "Dynamic audio reproduction with linear loudspeaker arrays," *J. Audio Eng. Soc*, vol. 67, no. 4, pp. 190–200, 2019.

[17] R. O. Duda and W. L. Martens, "Range dependence of the response of a spherical head model," *The Journal of the Acoustical Society of America*, vol. 104, no. 5, pp. 3048–3058, 1998.

[18] B. B. Bauer, "Phasor analysis of some stereophonic phenomena," *The Journal of the Acoustical Society of America*, vol. 33, no. 11, pp. 1536–1539, 1961.

[19] F. M. Fazi, M. Noisternig, and O. Warusfel, "Representation of sound fields for audio recording and reproduction," in *Acoustics*, 2012.

[20] O. Kirkeby, P. A. Nelson, H. Hamada, and F. Orduna-Bustamante, "Fast deconvolution of multichannel systems using regularization," *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 2, pp. 189–194, 03 1998.

[21] I.-R. BS775-1, "Multichannel sterephonic sound system with and without accompanying picture," International Telecommunications Union, Tech. Rep., 1994.

[22] E. G. Williams, *Sound Radiation and Nearfield Acoustical Holography*, 1st ed. London: Academic Press, 1999.

[23] M. Kolundzija, C. Faller, and M. Vetterli, "Sound field recording by measuring gradients," in *Audio Engineering Society Convention 128*, 2010.

[24] P. Morse and K. Ingard, *Theoretical Acoustics*, ser. International series in pure and applied physics. Princeton University Press, 1986.

[25] J.-M. Jin, *Theory and Computation of Electromagnetic Fields*. John Wiley and Sons, 2010.

[26] R. Mehrem, "The plane wave expansion, infinite integrals and identities involving spherical bessel functions," *Applied Mathematics and Computation*, vol. 217, no. 12, pp. 5360 – 5365, 2011.

[27] H. Møller, "Fundamentals of binaural technology," *Applied Acoustics*, vol. 36, no. 3, pp. 171–218, 1992.

[28] J. Meyer and G. Elko, "Spherical microphone array for spatial sound recording," in *Audio Engineering Society Convention 115*, 2003.

[29] ——, "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, 2002.

[30] J. Daniel, "Spatial sound encoding including near field effect: Introducing distance coding filters and a viable, new ambisonic format," in *Audio Engineering Society 23rd International Conference, Copenhagen, Denmark*, 2003.

Jacob Hollebon Jacob Hollebon is a Postdoctoral Researcher at the Institute of Sound and Vibration Research at the University of Southampton. Jacob graduated from the University of Warwick in 2017 with a BSc in Physics. He then completed the MSc in Acoustical Engineering at the University of Southampton in 2018, with a thesis on spatial audio reproduction for multiple listeners. In 2018, he joined the Virtual Acoustics and Audio Engineering team to begin a PhD in 3D audio reproduction methods over loudspeaker arrays. This research focused on developing novel signal processing techniques founded in existing 3D audio methods such as Sound Field Reproduction, Ambisonics and Crosstalk Cancellation. Jacob is the holder of the 2018 ISVR Elsevier prize, as well as being supported by two grants from the AES Educational Foundation, where he is also the 2018 Emil Torick Scholar.

Filippo Mari Fazi Filippo Maria Fazi is Professor of Acoustics and Signal Processing at the Institute of Sound and Vibration Research of the University of Southampton, where he also serves as head of the Acoustics Group and leads the Virtual Acoustics and Audio Engineering Team. His research interests include acoustics, audio technologies, electroacoustics and digital signal processing, with special focus on acoustical inverse problems, multi-channel systems (including Ambisonics and Wave Field Synthesis), virtual acoustics, and microphone arrays. He is the author of more than 150 scientific publications and several patents. Dr Fazi was awarded a research fellowship by the Royal Academy of Engineering (2010) and the Tyndall Medal by the Institute of Acoustics (2018). He is a fellow of the Audio Engineering Society, a member of the Institute of Acoustics and is co-founder and chief scientist at Audioscenic, a start-up company that develops and commercialises 3D audio and loudspeaker array technologies.