



**QUEEN'S
UNIVERSITY
BELFAST**

Binaural Reproduction of Finite Difference Simulations Using Spherical Array Processing

Sheaffer, J., Van Walstijn, M., Rafaely, B., & Kowalczyk, K. (2015). Binaural Reproduction of Finite Difference Simulations Using Spherical Array Processing. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 23(12), 2125-2135. Article 7194789. <https://doi.org/10.1109/TASLP.2015.2468066>

Published in:

IEEE/ACM Transactions on Audio, Speech, and Language Processing

Document Version:

Peer reviewed version

Queen's University Belfast - Research Portal:

[Link to publication record in Queen's University Belfast Research Portal](#)

Publisher rights

© 2015 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other users, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works for resale or redistribution to servers or lists, or reuse of any copyrighted components of this work in other works

General rights

Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

Open Access

This research has been made openly available by Queen's academics and its Open Research team. We would love to hear how access to this research benefits you. – Share your feedback with us: <http://go.qub.ac.uk/oa-feedback>

Binaural Reproduction of Finite Difference Simulations using Spherical Array Processing

Jonathan Sheaffer, Maarten van Walstijn, Boaz Rafaely, *Senior Member, IEEE*, and Konrad Kowalczyk

Abstract—Due to its efficiency and simplicity, the finite difference time domain method is becoming a popular choice for solving wideband, transient problems in various fields of acoustics. So far, the issue of extracting a binaural response from finite difference simulations has only been discussed in the context of embedding a listener geometry in the grid. In this paper we propose and study a method for binaural response rendering based on a spatial decomposition of the sound field. The finite difference grid is locally sampled using a volumetric array of receivers, from which a plane wave density function is computed and integrated with free-field head related transfer functions, in the spherical harmonics domain. The volumetric array is studied in terms of numerical robustness and spatial aliasing. Analytic formulas that predict the performance of the array are developed, facilitating spatial resolution analysis and numerical binaural response analysis for a number of finite difference schemes. Particular emphasis is placed on the effects of numerical dispersion on array processing and on the resulting binaural responses. Our method is compared to a binaural simulation based on the image method. Results indicate good spatial and temporal agreement between the two methods.

Index Terms—FDTD, simulation, finite difference methods, room acoustics, binaural processing, microphone arrays, sound reproduction

I. INTRODUCTION

WITH the growing availability of computing resources and recent advances in parallel processing, the Finite Difference Time Domain (FDTD) method is becoming a feasible choice for the simulation of room acoustics [1], environmental acoustics [2] and musical acoustics [3]. Approximating a solution to the wave equation in the discrete space/time domain, the FDTD method is advantageous for solving transient, wave dominated and broadband problems, many of which may benefit from *auralization* [4]. In the FDTD method, auralizations can be created by computing a binaural output from a grid excited by an audio signal, or by computing a binaural impulse response and convolving it with free-field recordings at a post-processing stage. To generalize, in this paper we shall simply refer to the binaural output of an FDTD simulation as a *binaural response*. Such a binaural response can be seen as a superposition of plane waves filtered by Head Related Transfer Functions (HRTFs), representing the

frequency and directional characteristics of the human head (and sometimes torso). In order to render binaural responses with FDTD simulations, the directional information - usually in the form of plane waves - must first be extracted from the grid variables.

One way to obtain a binaural response is by directly embedding a geometric model of a listener in the grid, in which case two receivers can be used, placed at the positions of the left and right ear canals. As a rough approximation, Murphy and Beeson [5] embedded a circular object in a 2D finite difference model and evaluated resulting interaural time difference (ITD) cues. This approach was further extended by Webb and Bilbao [6], as well as by Sheaffer et al. [7] who employed a full 3D model of a human head and, additionally, evaluated interaural level difference (ILD) cues. In these cases, resulting binaural cues were largely consistent with those experienced by the human auditory system. Nonetheless, modeling the fine geometric structure of the pinna [8], [9], [10] requires a considerably higher grid resolution than for modeling interaural cues. This imposes a significant computational burden and may also introduce errors when non-conformal boundary conditions are employed [11]. In addition, when embedding a listener in the grid, the transfer functions of the room and the head are jointly computed. Therefore, the entire simulation needs to be repeated when computing a sound field for different head-rotations or for personalized HRTFs. This places a practical limitation on using the method for applications such as motion-tracked binaural reproduction [12].

As an alternative to embedding a listener model in the grid, it is possible to render a binaural response by means of direct beamforming synthesis [13], or by extracting the plane-wave components of the soundfield [14] and spatially integrating them with pre-measured HRTFs. While the former approach is more computationally efficient, the latter provides freedom to manipulate the soundfield or the HRTFs at a post-processing stage. Both approaches have been discussed in the general context of grid-based simulation methods, but neither have been directly applied to FDTD as yet. Since the FDTD method inherently involves numerical dispersion, which is a frequency- and direction- dependent discrepancy in the phase velocities of the propagated waves [15], it is important to understand the role it plays in the synthesis of binaural responses.

The topic of auralizing FDTD data has also been addressed for loudspeaker reproduction, see e.g. [16], [17]. Of particular relevance here is a study by Southern et al. [18], who suggested to spatially encode an FDTD sound field

J. Sheaffer and B. Rafaely are with the Department of Electrical and Computer Engineering Ben-Gurion University of the Negev Beer-Sheva 84105, Israel. e-mail: sheaffer@ee.bgu.ac.il.

M. van Walstijn is with the Sonic Arts Research Centre (SARC) and the School of Electronics, Electrical Engineering, and Computer Science, Queens University, Belfast BT7 1NN, UK

K. Kowalczyk is with the Department of Electronics, AGH University of Science and Technology, 30-059 Krakow, Poland. e-mail: konrad.kowalczyk@agh.edu.pl.

in ambisonics format using a differential receiver array and demonstrated loudspeaker-based auralizations for 2D grids using array orders up to 3. However, the literature on finite difference modeling still does not offer a broadband and robust method for directly rendering 3D binaural responses using free-field HRTFs. In addition, the role of numerical dispersion in binaural response simulation has yet to be established.

In this paper we address the problem of modeling binaural receivers in FDTD using a plane-wave decomposition approach. We locally sample the soundfield using a volumetric array of receivers, which can be seen as a special case of the spherical shell microphone array [19]. By applying a spherical harmonics transform to the sampled data, we approximate a plane-wave density function from which a binaural response is computed. The contributions of this paper are as follows:

- 1) A method to render binaural responses from modeled FDTD data and pre-measured HRTFs is presented (Sec. III). A preliminary formulation of this method has been proposed in a recent conference publication by the authors [20]. This work is extended here through studying numerical robustness, as well as spatial aliasing for different FDTD design parameters (Sec. IV).
- 2) Analytic formulas to predict the effects of numerical dispersion on array directivity and on the resulting binaural response are developed and validated. Using these formulas, we study the effects of dispersion on spatial decomposition and evaluate the overall binaural reproduction error in different FDTD schemes (Sec. V).
- 3) For completeness, the applicability of the method to the rendering of binaural room impulse responses is demonstrated and validated by comparison with an image source model in a small rectilinear room with rigid walls (Sec. VI).

II. SOUND FIELD REPRESENTATION

In this section, the governing physical equations in continuous space and time are summarized. These form the basis of the discrete formulations further described in Sec. III.

A. Binaural Response Model

Consider a sound field governed by the homogeneous acoustic wave equation,

$$\nabla^2 p(\mathbf{r}, t) - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} p(\mathbf{r}, t) = 0, \quad (1)$$

where c is the speed of sound, ∇^2 is the Laplace operator and $p(\mathbf{r}, t)$ is the field variable, which is here assumed to be sound pressure. The wave equation can be described in Cartesian coordinates, in which case $\mathbf{r} \equiv (x, y, z) \in \mathbb{R}^3$, or in standard spherical coordinates, in which case $\mathbf{r} \equiv (r, \Omega)$, where r denotes radial distance and $\Omega \equiv (\theta, \phi) \in \mathbb{S}^2$ denotes direction in terms of elevation, θ , and azimuth ϕ . Fundamental solutions to (1) may have spherical symmetry; however, once waves have propagated a sufficiently long distance, the resulting sound field can be seen as a continuum of plane waves, described by a plane wave density function, $a(k, \Omega)$, where $k = 2\pi f/c$ is the wavenumber, and f is the frequency. Given

a sound field composed of plane waves, a binaural response can be synthesized by convolution with free-field HRTFs, as follows:

$$p^l(k) = \int_{\Omega \in \mathbb{S}^2} a(k, \Omega) H^l(k, \Omega) d\Omega, \quad (2)$$

where $H^l(k, \Omega)$ is an HRTF catalog for the left-ear, $p^l(k)$ is the sound pressure at the left ear, and $\int d\Omega = \int \sin \theta d\theta d\phi$. The integration in (2) involves a plane wave density function; therefore, a method is required to estimate $a(k, \Omega)$ from the field variables. In the following section, we propose a method to extract $a(k, \Omega)$ from pressure signals; however, similar techniques can be developed for other field variables occasionally used in FDTD simulations, such as particle velocity or a velocity potential.

B. Expansion in Spherical Harmonics

Assuming that $a(k, \Omega)$ is square-integrable over Ω , its spherical Fourier transform, denoted by $a_{nm}(k, r)$, and the corresponding inverse transform are given by [21]

$$a_{nm}(k) = \int_{\Omega \in \mathbb{S}^2} a(k, \Omega) [Y_n^m(\Omega)]^* d\Omega, \quad (3)$$

$$a(k, \Omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^n a_{nm}(k) Y_n^m(\Omega), \quad (4)$$

where the operator $[\cdot]^*$ denotes complex conjugation. The spherical harmonics function, $Y_n^m(\cdot)$, is given by [21]

$$Y_n^m(\Omega) = \sqrt{\frac{(2n+1)(n-m)!}{4\pi(n+m)!}} P_n^m(\cos \theta) e^{im\phi}, \quad (5)$$

where n and m denote order and degree, respectively, and $P_n^m(\cdot)$ is the associated Legendre function. In a similar fashion, $p_{nm}(k, r)$ denotes the spherical Fourier transform of $p(k, r, \Omega)$, which is the representation of $p(\mathbf{r}, t)$ in the frequency domain. Since $p_{nm}(k, r)$ is related to $a_{nm}(k)$ through a radial function [22], [19], the following expression holds:

$$\begin{aligned} p_{nm}(k, r) &= b_n(kr) \int_{\Omega \in \mathbb{S}^2} a(k, \Omega) [Y_n^m(\Omega)]^* d\Omega \\ &= b_n(kr) a_{nm}(k), \end{aligned} \quad (6)$$

where $b_n(kr)$ is the radial function, which is given for an open-sphere (i.e. with no scattering objects) by $b_n(kr) = 4\pi i^n j_n(kr)$ [21], $j_n(\cdot)$ is the n^{th} order spherical Bessel function and $i = \sqrt{-1}$. Accordingly, it can be shown that the relationship between sound pressure and the plane wave density function is

$$p(k, r, \Omega) = \sum_{n=0}^{\infty} b_n(kr) \sum_{m=-n}^n a_{nm}(k) Y_n^m(\Omega). \quad (7)$$

Once the function $a_{nm}(k)$ is known, a binaural signal can be rendered by employing the spherical harmonics transform of (2), [23]:

$$p^l(k) = \sum_{n=0}^{\infty} \sum_{m=-n}^n \tilde{a}_{nm}^*(k) H_{nm}^l(k), \quad (8)$$

where

$$\tilde{a}_{nm}(k) = (-1)^m [a_{n(-m)}(k)]^* \quad (9)$$

is the representation of $a^*(k, \Omega)$ in the spherical harmonics (SH) domain. The transfer function for the right ear can be computed in a similar fashion with the right ear HRTF, $H^r(k, \Omega)$. In order to approximate a binaural response from modeled pressure signals, one needs to solve (7) for $a_{nm}(k)$, which then allows using (8) to compute the sound pressure at the ear directly in the SH domain. Performing this process with FDTD data is discussed in detail in Sec. III.

C. Spatial Resolution

The formulation described in Sec. II-B can be seen as a two-step process. First, the sound field is decomposed into its plane wave components [22], followed by binaural rendering in the SH domain [23]. Such a *Plane Wave Decomposition* (PWD) process is akin to employing a maximum-directivity beamformer whose look directions, $\Omega_L \equiv (\theta_L, \phi_L)$, are tuned to the directions of the arriving plane waves [22]. Accordingly, the directivity of the beamformer controls the spatial resolution with which the sound field is decomposed. In the SH domain, a sound field consisting of a single plane wave incident at Ω_0 has the density function $a_{nm} = [Y_n^m(\Omega_0)]^*$; therefore, from (6), the directivity of an ideal PWD beamformer can be described by

$$\begin{aligned} y(\Omega_0, \Omega_L) &= \sum_{n=0}^{\infty} \sum_{m=-n}^n \frac{p_{nm}(k, r)}{b_n(kr)} Y_n^m(\Omega_L) \\ &= \sum_{n=0}^{\infty} \sum_{m=-n}^n \frac{b_n(kr) [Y_n^m(\Omega_0)]^*}{b_n(kr)} Y_n^m(\Omega_L) \\ &= \delta(\cos \theta_L - \cos \theta_0) \delta(\phi_L - \phi_0), \end{aligned} \quad (10)$$

where $y(\Omega_0, \Omega_L)$ is the beamformer's output. If (10) is truncated at some finite order N , then it follows from the spherical harmonics addition theorem and the Christoffel summation formula that (10) reduces to

$$y(\Omega_0, \Omega_L) = \frac{N+1}{4\pi(\cos \Theta - 1)} [P_{N+1}(\cos \Theta) - P_N(\cos \Theta)], \quad (11)$$

where Θ is the angle between Ω_0 and Ω_L . These expressions are the basis of the numerical directivity formulations further developed in Sec. V-B.

III. NUMERICAL FORMULATION

In this section, a numerical formulation of binaural response rendering that is directly applicable to FDTD is proposed. To simulate wave propagation using the FDTD method, the sound field is discretized on a Cartesian grid such that $(x, y, z, t) \rightarrow [dX, fX, gX, uT]$, where u and $[d, f, g]$ are the index positions in discrete time and space, respectively, and X and T are the spatial and temporal sample periods. Correspondingly, the wave equation, (1), can be modeled as [1]

$$[\delta_t^2 - \lambda^2(\delta_x^2 + \delta_y^2 + \delta_z^2 + C_1 \delta_x^2 \delta_y^2 + C_1 \delta_x^2 \delta_z^2 + C_1 \delta_y^2 \delta_z^2 + C_2 \delta_x^2 \delta_y^2 \delta_z^2)] p|_{d,f,g}^u = 0, \quad (12)$$

with δ_D^2 denoting a second-order finite difference operator over the dimension D ; for example,

$$\delta_x^2 p|_{d,f,g}^u \equiv p|_{d+1,f,g}^u - 2p|_{d,f,g}^u + p|_{d-1,f,g}^u. \quad (13)$$

A detailed discussion on finite difference operators can be found in recent literature [3]. The Courant number, $\lambda = cT/X$, and constants C_1 and C_2 , are chosen according to the desired *finite difference scheme*, see e.g. Table I in [1]. Consider now signals captured at Q grid nodes that are arbitrarily distributed around some point, each having a radial distance, $r_q = \tilde{r}_q X$, and angle, $\Omega_q \equiv (\theta_q, \phi_q)$, with respect to that point; \tilde{r}_q is measured in nodes and the spatial period X is here employed as a normalization constant. The pressure at each of the receivers is captured and transformed to the frequency domain, resulting in a frequency-dependent vector,

$$\mathbf{p} = [p_1(k), p_2(k), \dots, p_Q(k)]^T, \quad (14)$$

where $(\cdot)^T$ denotes transposition. Following (7), the pressure at each receiving node can be approximated by [19]

$$p(k, \tilde{r}_q, \Omega_q) \approx \sum_{n=0}^N \sum_{m=-n}^n a_{nm}(k) b_n(k \tilde{r}_q X) Y_n^m(\Omega_q), \quad (15)$$

with $1 \leq q \leq Q$. The approximation becomes an equality as $N \rightarrow \infty$ or if the sound field is known to be order-limited at some finite order N . This relationship can be expressed in matrix form as follows:

$$\mathbf{p} = \mathbf{B} \mathbf{a}_{nm}, \quad (16)$$

where \mathbf{a}_{nm} is a $(N+1)^2 \times 1$ vector representing the plane wave coefficients of the sound field,

$$\mathbf{a}_{nm} = [a_{00}, a_{1(-1)}, a_{10}, a_{11}, \dots, a_{NN}]^T, \quad (17)$$

and the $Q \times (N+1)^2$ matrix \mathbf{B} is given by

$$\mathbf{B}^T = \begin{bmatrix} b_0(k \tilde{r}_1 X) Y_0^0(\Omega_1) & \dots & b_0(k \tilde{r}_Q X) Y_0^0(\Omega_Q) \\ b_1(k \tilde{r}_1 X) Y_1^{-1}(\Omega_1) & \dots & b_1(k \tilde{r}_Q X) Y_1^{-1}(\Omega_Q) \\ \vdots & \ddots & \vdots \\ b_N(k \tilde{r}_1 X) Y_N^N(\Omega_1) & \dots & b_N(k \tilde{r}_Q X) Y_N^N(\Omega_Q) \end{bmatrix}. \quad (18)$$

The plane wave density function can be approximated from the pressure signals by

$$a_{nm}(k) \approx \sum_{q=1}^Q c_{nm}^q(k) p(k, \tilde{r}_q, \Omega_q), \quad (19)$$

where c_{nm}^q are quadrature coefficients transforming sound pressure into a plane wave density function. If the sound field is sampled such that $Q \geq (N+1)^2$, where N is the highest SH order to be decomposed, then (16) has a solution in a least-squares sense [19]. The PWD can then be approximated numerically as follows:

$$\mathbf{a}_{nm} \approx \mathbf{C} \mathbf{p} = \mathbf{B}^\dagger \mathbf{p}, \quad (20)$$

where $\mathbf{C} = \mathbf{B}^\dagger$ is a $(N+1)^2 \times Q$ matrix holding the coefficients c_{nm}^q on row q and column $n^2 + n + m + 1$. The matrix $\mathbf{B}^\dagger = (\mathbf{B}^H \mathbf{B})^{-1} \mathbf{B}^H$ is the Moore-Penrose pseudoinverse of \mathbf{B} , and $(\cdot)^H$ denotes the conjugate transpose. All

matrices are frequency dependent. The notion of employing quadrature coefficients to perform a numerical PWD, and hence the definition of the matrix \mathbf{C} , are here introduced for mathematical convenience and shall be further referred to in Sec. V-B.

An important point is that the magnitude of the radial function, $b_n(k\tilde{r}X)$, vanishes as $(k\tilde{r}X) \rightarrow 0$ for $n > 0$. This, in turn, has a direct effect on the numerical robustness of the system. To ensure that \mathbf{B} is well-conditioned, we opt to regularize the problem by employing soft-limited radial filters, which have been shown to be nearly free of spatial and temporal artifacts [24], [25]. In this case it is possible to construct a regularized matrix $\tilde{\mathbf{B}}$ by substituting (21) into (18). With some notational changes, the modified radial functions are given as follows [24]:

$$\tilde{b}_n(k\tilde{r}X) = \frac{i^n j_n(k\tilde{r}X)}{2\alpha\Gamma(k\tilde{r}X)|i^n j_n(k\tilde{r}X)|}, \quad (21)$$

where $\alpha = 10^{(\Delta/20)}$, Δ is the total allowable attenuation of $b_n(k\tilde{r}X)$ in decibels and $\Gamma(k\tilde{r}X) = \arctan(1/(8\alpha|i^n j_n(k\tilde{r}X)|))$. In practice, such regularization trades off numerical robustness with spatial resolution, as it attenuates high-order SH components.

IV. ARRAY DESIGN AND ANALYSIS

Some of the specific properties of the array, namely the total number of receivers and their spatial distribution, have a direct effect on the accuracy of PWD and, hence, on the resulting binaural signals. The array processing literature portrays a wide range of *spatial sampling schemes*, as well as *array configurations*, which describe the types of sensors and properties of the array surfaces [26]. One array configuration particularly relevant to this study is the spherical shell array [19], in which sound pressure is sampled within a volume confined by two concentric spheres. Using this type of array, sampling can be performed using omnidirectional (pressure) sensors in an open-sphere configuration, while avoiding robustness issues associated with the zeros of the open-sphere radial functions [19]. The spherical shell array also provides significant design freedom, as receiver positions can depend on both angle and radius and do not need to conform to a pre-determined sampling scheme. In the context of FDTD modeling, this has the convenience of allowing one to place receivers directly at grid nodes, thus avoiding the need to perform spatial interpolation.

In the original spherical shell design, a genetic algorithm was employed for minimizing the number of required spatial samples [19]. Unlike real microphone arrays, such minimization is not critical in a numerical model, as the array size is only constrained by the available system memory. Additionally, sampling the sound field at a large number of nodes, i.e. $Q > (N+1)^2$, results in an over-determined system, which is beneficial for increasing its robustness [19] and, as will be discussed shortly, also for decreasing its associated aliasing error. Therefore, in this study we opt to include all spatial samples confined within a spherical volume of radius $r = \tilde{r}X$, as shown in Figure 1. This array design will be

further referred to as a Full-Volume Spherical (FVS) array. In more memory critical applications (such as processing on a GPU), one can utilize the techniques reported in [19] to further reduce memory requirements.

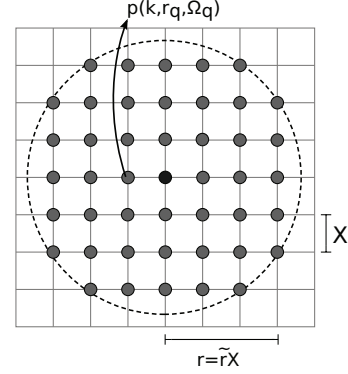


Fig. 1. A 2D section of a volumetric spherical array. The array radius \tilde{r} and the FDTD spatial period, X , control the total number of the receivers in the array.

For the specific application of binaural simulation, it is desired that the array would perform well in broadband. The spectral bandwidth of the associated PWD is normally limited at low-frequencies, by numerical robustness, and at high-frequencies, by spatial aliasing. Accordingly, it is useful to study the effect of the parameters X and \tilde{r} on the overall performance of the array. These considerations are discussed in detail in the remainder of this section.

A. Numerical Robustness

Due to the low magnitude of the spherical Bessel function at low frequencies and high orders, the matrix \mathbf{B} may become ill conditioned, meaning that any errors present in \mathbf{p} will be significantly amplified. In an FDTD model, such errors can be attributed to numerical dispersion, as well as to the finite computation precision. This may appear to be counter-intuitive, as numerical dispersion vanishes as $k \rightarrow 0$ and precision errors are very low to begin with. However, as will be further demonstrated in Sec. V, the matrix inversion related to spherical array processing may indeed give rise to such errors when \mathbf{B} is not appropriately regularized.

One way to evaluate numerical robustness is by computing the *condition number* of the matrix \mathbf{B} [19]. Figure 2 shows the 2-norm condition number, $\kappa(\mathbf{B})$, for a full volumetric spherical array of $\tilde{r} = 10$ nodes designed at $N = 12$, without regularization (FVS) and a full volumetric spherical array with regularization (FVS-R), corresponding to $\Delta = 40$ dB. For comparison purposes, two single sphere arrays of $r = 0.1$ m (equivalent to $\tilde{r} = 10$ nodes at $X = 0.01$ m) are designed using a Lebedev sampling scheme (230 points, corresponding approximately to $N = 12$) [26]. The open sphere array (OS) is an example of a poor broadband design, as the condition number peaks at frequencies corresponding to zeros of the spherical Bessel function. In contrast, the rigid sphere array (RS) avoids these artifacts, thus providing an ideal reference in terms of robustness [26], albeit at the price of introducing a scattering object into the sound field.

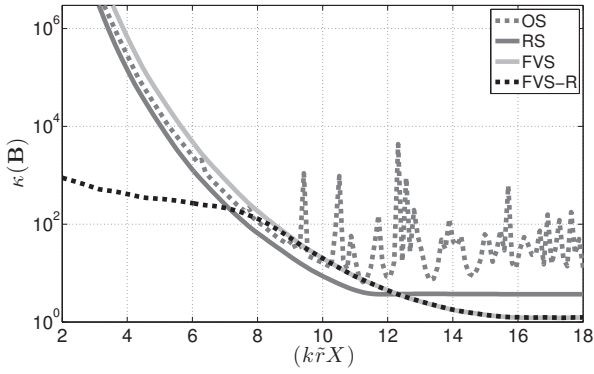


Fig. 2. Condition number of a matrix \mathbf{B} as a function of $(k\tilde{r}X)$ shown for a number of $N = 12$ arrays. OS: Open sphere, RS: Rigid sphere, FVS: Full volumetric sphere, FVS-R: Regularized full volumetric sphere.

In all arrays, the condition number rapidly decreases as kr approaches zero. However, with the aid of regularization, robustness can be significantly improved, even at low frequencies, as seen in the FVS-R curve. It is also worthwhile noting that since $\tilde{r} = r/X$, these results are directly scalable to different combinations of sample periods and radii. In summary, it can be said that the FVS design features a numerical robustness comparable to that of a RS array, yet at the same time maintains a sound field free of artificial scattering.

B. Spatial Aliasing

While numerical robustness limits the bandwidth of the FVS array at low frequencies, it may be constrained at high frequencies by spatial aliasing. In non-volumetric configurations, a spherical array is normally designed such that the highest frequency of interest complies with $kr \leq N$, where r is the radius of the sphere [27]. The FVS array, however, does not feature a single radius and, therefore, the contribution of spatial aliasing needs to be evaluated numerically. Drawing from [27], [28], we consider an array designed to operate up to some arbitrary order N . The goal of the following analysis is to quantify the amount of spatial aliasing that will contaminate the array output, based on the structure of the array. For a sound field limited to an order N , and sampled such that $Q \geq (N+1)^2$, one can expect that $\mathbf{B}^\dagger \mathbf{B} = \mathbf{I}$, where \mathbf{I} is the identity matrix. Consider now the same array at some higher sound field order, \hat{N} , resulting in a system characterized by $\mathbf{p} = \hat{\mathbf{B}} \hat{\mathbf{a}}_{\text{nm}}$. To quantify aliasing error in the process of computing a PWD for the equations related to \hat{N} , we further define a matrix $\mathbf{D} = \mathbf{B}^\dagger \hat{\mathbf{B}}$, such that

$$\mathbf{D} = (\mathbf{B}^H \mathbf{B})^{-1} \mathbf{B}^H \hat{\mathbf{B}}. \quad (22)$$

For the special case of $N = \hat{N}$, we expect \mathbf{D} to reduce to the identity matrix. Accordingly, for $N < \hat{N}$, the matrix \mathbf{D} will also contain the contribution of spatial aliasing in the frequency range $N < (k\tilde{r}X) \leq \hat{N}$. Setting \hat{N} to some very high (yet finite) order, the total contribution of spatial aliasing can then be quantified as

$$\epsilon = \|\mathbf{D} - \mathbf{I}\|_2, \quad (23)$$

where \mathbf{I} may be a non-square (zero padded) version of the unit matrix corresponding to the dimensions of \mathbf{D} .

To study the effects of modifying the FDTD sample period, the aliasing error, ϵ , was calculated at $N = 12$ for FVS arrays of radii $\tilde{r} = 5, 7$ and 10 nodes. For reference, a single RS array of radius $r = 0.1\text{m}$ was designed using a Lebedev scheme matching $N = 12$ (230 samples only on the surface of the sphere). Figure 3 shows the aliasing errors for these four cases.

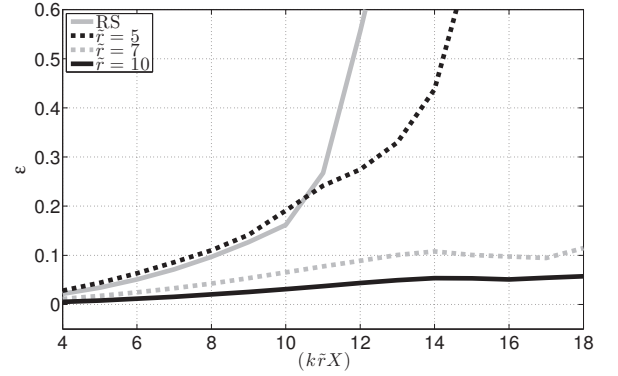


Fig. 3. Aliasing error, ϵ , for three $N = 12$ FVS arrays of radius \tilde{r} (in nodes). The reference is a rigid sphere (RS) array, designed with an $N = 12$ Lebedev scheme, $r = 0.1\text{m}$ corresponding to $\tilde{r} = 5, 6$ and 10 nodes, at $X = 0.02\text{m}$, 0.0143m and 0.01m , respectively.

For the RS array, the aliasing error rapidly increases with kr , where at $(k\tilde{r}X) = N$ the error is about $\epsilon = 0.5$. For the $\tilde{r} = 5$ nodes FVS array, the aliasing is comparable to that of the RS reference up to $(k\tilde{r}X) \approx N$, and is slightly lower than the RS at higher spatial frequencies. In contrast, the $\tilde{r} = 7$ and $\tilde{r} = 10$ arrays feature considerably lower aliasing even for $(k\tilde{r}X) \geq N$, and remain within $\epsilon \leq 0.1$ even up to $(k\tilde{r}X) = 1.5N$. This serves to emphasize the significance of a substantially oversampled array, as is achieved by the concept of the FVS array.

C. Array Size

For a volumetric spherical array defined over a rectilinear grid, the total number of nodes is approximately $(4/3)\pi r^3$. Since there exists no pre-defined sampling scheme for a volumetric array, it is useful to propose a rule of thumb, relating the array volume to a desired order and operating frequency. Assuming an array designed to operate ideally at $(k\tilde{r}X) = N$, the total number of nodes would be:

$$Q \approx \frac{1}{6\pi^2} \left(\frac{N\Delta}{f_c} \right)^3, \quad (24)$$

where f_c is the normalized lower cutoff frequency of the array. This results in a significantly oversampled array, ensuring a robust decomposition with minimal aliasing. In order to reduce the computational cost of array processing, the number of the array nodes can be further reduced by confining the array from a full volumetric shape to a shell defined by two concentric spheres with a radii ratio of β , resulting in

$$Q \approx \frac{1}{6\pi^2} \left(1 - \frac{1}{\beta^3} \right) \left(\frac{N\Delta}{f_c} \right)^3. \quad (25)$$

As a rule-of-thumb, a value of $\beta = 1.2$ is near-optimal for a wide range of arrays [29]. For example, designing the $N = 12$ array described in preceding sections using this rule of thumb, results in a reduction of over 50% in the total number of nodes, with a relatively small increase of the condition number at $\tilde{k}rx = N$, ($\kappa = 4.346$ compared to $\kappa = 3.666$), but a more significant increase in aliasing error at $\tilde{k}rx = N$, ($\epsilon = 0.12$ compared to $\epsilon = 0.043$). In some cases, it may be possible to further reduce the total number of array nodes down to the lower bound of $Q = (N + 1)^2$, by selecting samples equally spaced on the radial dimension [19]. This process would, however, require numerical optimization and possibly also performing spatial interpolation over the grid nodes.

V. THE EFFECTS OF DISPERSION

A. Mathematical Formulation

A fundamental drawback of the FDTD method is numerical dispersion, which contaminates modeled signals with frequency- and direction- dependent errors. For the general family of compact explicit FDTD schemes used in this study, the dispersion relation is given by [1]

$$\sin^2(\pi fT) = \lambda^2[(s_x + s_y + s_z) - 4C_1(s_x s_y + s_x s_z + s_y s_z) + 16C_2 s_x s_y s_z], \quad (26)$$

with

$$\begin{aligned} s_x &= \sin^2\left(\frac{1}{2}\tilde{k}X \cos\phi \sin\theta\right), \\ s_y &= \sin^2\left(\frac{1}{2}\tilde{k}X \sin\phi \sin\theta\right), \\ s_z &= \sin^2\left(\frac{1}{2}\tilde{k}X \sin\theta\right), \end{aligned} \quad (27)$$

where $\tilde{k} \equiv \tilde{k}(\theta, \phi, f)$ is the numerical wavenumber in the direction (θ, ϕ) at frequency f . It is also worthwhile noting that if a nearly-isotropic FDTD scheme is used, then it is possible to employ frequency warping to further reduce the overall effects of dispersion [30]. This can be modeled using a post-warped numerical wavenumber, $\tilde{k}_w = \tilde{k}\tilde{c}_w/c$, where \tilde{c}_w is the numerical wave velocity in either the diagonal or the axial directions. Accordingly, in this paper such post-warping is applied to all results involving the *Interpolated Isotropic* (IISO2) finite difference scheme, with \tilde{c}_w chosen in the axial direction.

To better understand the possible effects of dispersion, we consider a single unit amplitude plane wave propagating in a dispersive medium and impinging on the surface of an open sphere of radius r . At this point a continuous sphere is assumed, providing a focus on dispersion without the effects of spatial sampling. Since the sound field is composed of a single plane wave, it can be said that the numerical wave propagation velocity \tilde{c} is homogeneous across the entire physical domain (although it differs from its analytic counterpart, c). Accordingly, it is possible to express the pressure on the surface of the sphere as follows:

$$p_{nm}(\tilde{k}) = b_n(\tilde{k}r) [Y_n^m(\Omega_0)]^*, \quad (28)$$

where $\Omega_0 \equiv (\theta_0, \phi_0)$ is the plane wave angle of incidence and \tilde{k} is the numerical wavenumber which can be computed

from (26). Substituting $p_{nm}(\tilde{k})$ into the numerator of (10), the output of a PWD beamformer in a dispersive medium becomes,

$$\begin{aligned} \tilde{y}(k, \Omega_L) &= \sum_{n=0}^{\infty} \sum_{m=-n}^n \frac{p_{nm}(\tilde{k}, r)}{b_n(kr)} Y_n^m(\Omega_L) \\ &= \sum_{n=0}^{\infty} \sum_{m=-n}^n \underbrace{\frac{b_n(\tilde{k}r)}{b_n(kr)}}_{w_q} [Y_n^m(\Omega_0)]^* Y_n^m(\Omega_L). \end{aligned} \quad (29)$$

Compared to (10), the under-braced term in (29) cannot cancel out, as the radial functions for the analytic and numerical wavenumbers are different in most cases. Accordingly, a spatial delta function is fully recovered in a dispersive medium only at directions and frequencies for which $\tilde{k} = k$. We therefore expect the behaviour of the PWD beamformer to be dependent on the chosen finite difference scheme.

B. Numerical Directivity Analysis

To further study how PWD is affected by the FDTD design parameters, it is useful to develop a closed-form representation of the directivity of an FVS array embedded in a finite difference grid. This mathematical formulation will be further referred to as a *Numerical Directivity Analysis* (NDA). Unlike (29), which was formulated for a continuous sphere, the output of a sampled, order-limited PWD beamformer can be written as follows [19]:

$$y(k, \Omega_L) = \frac{4\pi}{(N+1)^2} \sum_{n=0}^N \sum_{m=-n}^n a_{nm}(k) Y_n^m(\Omega_L). \quad (30)$$

For a plane wave propagating in an FDTD grid, the pressure at the q^{th} receiver can be calculated using

$$p(\tilde{k}, \tilde{r}_q, \Omega_q, \Omega_0) = e^{i\tilde{k}\tilde{r}_q X \cos(\Theta_q)}, \quad (31)$$

where Θ_q is the angle between Ω_q and Ω_0 . Substituting (31) into (19), the plane wave density function in an FDTD grid is given by

$$a_{nm}(\tilde{k}) \approx \sum_{q=1}^Q c_{nm}^q(k) e^{j\tilde{k}\tilde{r}_q X \cos(\Theta_q)}, \quad (32)$$

Finally, substituting $a_{nm}(\tilde{k})$ into (30), the following NDA formula is obtained:

$$\begin{aligned} y(\tilde{k}, \Omega_L) &= \sum_{q=1}^Q e^{i\tilde{k}\tilde{r}_q X \cos \Theta_q} \\ &\times \underbrace{\frac{4\pi}{(N+1)^2} \sum_{n=0}^N \sum_{m=-n}^n c_{nm}^q(k) Y_n^m(\Omega_L)}_{w_q}. \end{aligned} \quad (33)$$

Observe that the weighting term w_q in (33) is dependent on k but independent of \tilde{k} , thus indicating that numerical dispersion manifests itself as an input error. Also note that, unlike ideal beam patterns, here, directivity is formulated as a function of the beamformer's look direction and, due to the existence of dispersion, also as a function of \tilde{k} . Since the values of c_{nm}^q

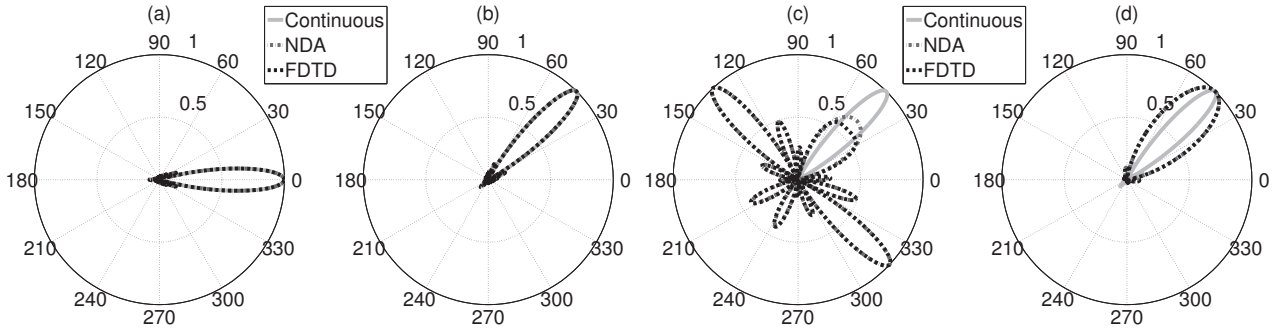


Fig. 4. Directivity patterns of a 12^{th} order array at $k\tilde{r}X = N$ for waves incident at (a) axial and (b) side-diagonal directions. For the same array, results are shown at the side-diagonal direction for (c) $k\tilde{r}X = 4$ and (d) $kr = 4$, with radial filter limiting of $\Delta = 10\text{dB}$. FDTD - finite difference simulation, NDA - numerical directivity analysis, Continuous - closed form solution. All data are normalized for visual clarity.

are generally unknown, it is convenient to solve (33) in matrix form. Defining a vector,

$$\mathbf{Y} = [Y_0^0(\Omega_L), Y_1^{-1}(\Omega_L), \dots, Y_N^N(\Omega_L)]^T \quad (34)$$

denoting the array's look direction, the weighting term can be written as $\mathbf{w} = \mathbf{C}\mathbf{Y}$, where \mathbf{C} is as defined in Sec. III. Accordingly, (33) can be computed as follows:

$$y = \left[\frac{4\pi}{(N+1)^2} \mathbf{B}^\dagger \mathbf{Y} \right]^T \tilde{\mathbf{p}}(\tilde{k}), \quad (35)$$

where

$$\tilde{\mathbf{p}}(\tilde{k}) = [e^{i\tilde{k}\tilde{r}_1 X \cos(\Theta_1)}, \dots, e^{i\tilde{k}\tilde{r}_q X \cos(\Theta_q)}]^T. \quad (36)$$

To validate the NDA formula, an FDTD simulation was executed using the Interpolated Wideband (IWB) scheme ($X = 0.01\text{m}$) for a source situated 150 nodes (1.5m) from the center of an FVS array of radius $\tilde{r} = 10$ nodes. The simulation was repeated for axial ($\Omega_0 \equiv (\pi/2, 0)$) and side-diagonal ($\Omega_0 \equiv (\pi/2, \pi/4)$) source incidence angles. For comparison with an ideal PWD beamformer, results were also generated in closed form using (11). To exclude any artefacts caused by spatial sampling, all arrays are studied below their aliasing limit.

Figures 4(a) and 4(b) show directivity patterns for a 12^{th} order array at $k\tilde{r}X = N$. Good visual agreement between the FDTD simulation and the NDA formula can be seen, indicating that the effects of dispersion on directivity are negligible at the studied frequency. This is attributed to the high numerical robustness at $k\tilde{r}X = N$ (see Fig. 2). For comparison, Figure 4(c) shows results for the same array at $k\tilde{r}X = 4$, where the numerical robustness is considerably lower. It is evident that the directivity patterns of both the FDTD and the NDA arrays no longer feature a dominant main lobe. Since all arrays are evaluated below their aliasing limit, differences between the NDA/FDTD results and the ideal (continuous) beamformer can be attributed only to numerical dispersion or to numerical precision errors. In double-precision arithmetic, precision errors are in the order of minus hundreds of decibels. Therefore, it can be postulated that the differences in beam patterns are caused by amplification of numerical dispersion, even though it is relatively low at the corresponding

frequency ($f/f_s = 0.065$). This was verified by computing (33) with an ideal wavenumber and comparing to the results of (10). Figure 4(d) shows results for $k\tilde{r}X = 4$, with a radial function limit of $\Delta = 10\text{dB}$. It can be seen that, due to the improved robustness, the main lobe of both the FDTD and NDA arrays point to the look direction. However, as previously suggested, this comes at the expense of a lower spatial resolution, which is evident in the increased width of the beamformer's main lobe.

C. Numerical HRTF Analysis

It is possible to further extend the NDA formula to predict the final sound pressure at the ear. Such a procedure, herein termed a *Numerical HRTF Analysis* (NHA), is useful for investigating the effects of dispersion on binaural response synthesis without needing to execute lengthy FDTD simulations. In particular, this is beneficial for studying the spatial effects of dispersion, in which case one needs to take samples from a relatively large number of incident directions. Similar to the NDA, we consider a single plane wave with a numerical propagation wavenumber, \tilde{k} , incident at Ω_0 . Following (8), the pressure at the left ear is

$$p^l(k) = \sum_{n=0}^N \sum_{m=-n}^n \tilde{a}_{nm}^*(\tilde{k}) H_{nm}^l(k). \quad (37)$$

Note the difference in \tilde{k} and k for the plane wave density and HRTF terms, respectively. Substituting (32) into (37) and considering the SH permutation defined in (9), the pressure at the left ear can be written as follows:

$$\begin{aligned} p^l(k) &= \sum_{n=0}^N \sum_{m=-n}^n \sum_{q=1}^Q \tilde{c}_{nm}^q(k) e^{j\tilde{k}\tilde{r}_q X \cos(\Theta_q)} H_{nm}^l(k) \\ &= \sum_{q=1}^Q e^{j\tilde{k}\tilde{r}_q X \cos(\Theta_q)} \sum_{n=0}^N \sum_{m=-n}^n (-1)^m c_{n(-m)}^q(k) H_{nm}^l(k). \end{aligned} \quad (38)$$

Defining now a vector denoting the SH coefficients for the left-ear HRTF at wavenumber k ,

$$\mathbf{h}^l(k) = [H_{00}^l(k), H_{1(-1)}^l(k), \dots, H_{NN}^l(k)]^T, \quad (39)$$

Eqn. (38) can be expressed in matrix form as follows:

$$p^l(k) = [\mathbf{B}^\dagger \mathbf{R} \mathbf{h}^l(k)]^T \mathbf{p}(\tilde{k}), \quad (40)$$

where \mathbf{R} is a $(N+1)^2 \times (N+1)^2$ permutation matrix converting $c_{nm}^q(k)$ into $\tilde{c}_{nm}^q(k)$ through (9).

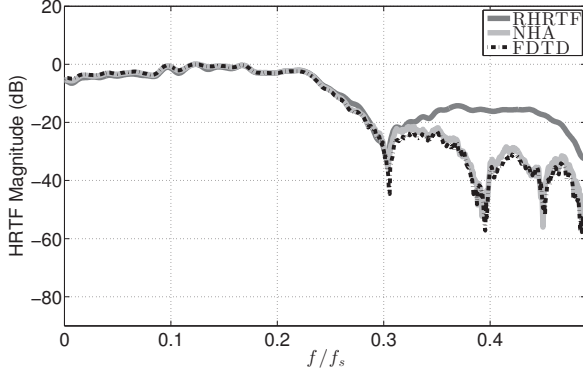


Fig. 5. Magnitude of modeled HRTFs as a function of normalized frequency. RHRTF - Reduced order HRTF (reference), NHA - Numerical HRTF analysis, FDTD - HRTF modeled in a finite difference simulation. The finite difference scheme is IWB, $f_s = 34350\text{Hz}$, $X = 0.01\text{m}$, $\tilde{r} = 10$ nodes, $N = 12$.

Figure 5 shows the left-ear HRTF magnitude for a single diagonally-incident wave, modeled in an FDTD simulation with the same parameters as in Sec. V-B (dashed black line). The simulation result is plotted against a numerical HRTF analysis (NHA - light gray line) and against a free-field HRTF measurement reduced to the SH order of the sound field decomposition, $N = 12$ (RHRTF - dark gray line). It can be seen that the HRTF and NHA curves are generally in agreement throughout the entire frequency spectrum, providing experimental validation of the NHA formula. All curves are in agreement up to $f/f_s = 0.3$, where the RHRTF curve begins to diverge due to the effects of dispersion, which are the strongest in the diagonal direction for the employed IWB scheme. For the sample rate of $f_s = 34350\text{Hz}$ this corresponds to about 10kHz, which is considerably higher than other objective and subjective dispersion limits reported in the literature [1], [31]. Perceptually, the artificial spectral notches at $f/f_s > 0.3$ may introduce conflicting monaural cues; these are directly related to sound localization in elevation and to the ability to resolve front-back confusion [32].

In Figure 6, an NHA was employed to predict modeled HRTFs at 338 incidence directions, spherically distributed around the listener in an equi-angle distribution. Each 3D plot shows the HRTF magnitude as a function of the incidence direction for a single frequency. For reference, an NHA was computed with an ideal wavenumber, k , hence ensuring that differences between the plots can occur only due to numerical dispersion. None of the computations employ regularization of the matrix \mathbf{B} . At $(k\tilde{r}X) = 1$, numerical robustness is extremely poor. The radial function $b_{1,2}(1)$ indicates amplification of over 118dB and hence gives rise to numerical dispersion errors, even though they are relatively low at the corresponding normalized frequency, e.g. $f/f_s = 0.0159$ for the IWB scheme. Note the visual difference between the reference and the IWB and SRL cases, showing that errors

TABLE I
HRTF REPRODUCTION ERROR (IN DECIBELS) FOR A VOLUMETRIC ARRAY OF $\tilde{r} = 10$ NODES DESIGNED ON A GRID OF $X = 0.01\text{m}$.

$(k\tilde{r}X)$	1	$N/4$	$N/2$	N	$1.5N$	$2N$	$2.5N$
Non-regularized							
IWB	38.9	7.2	-4.3	-13.8	-7.1	-2.2	0.3
SRL	41.9	9.4	-1.7	-11.3	-4.1	-0.3	0.5
IISO	41.5	12	-9.3	-12.7	-13.3	-3.9	26.3
Regularized, $\Delta = 10\text{dB}$							
IWB	-17.5	-15.3	-19.2	-14.3	-5.1	-0.6	0.1
SRL	-18.5	-14.7	-17.1	-11.9	-4.3	-0.2	0.5
IISO	-17.4	-17.7	-15.3	-15.3	-7.6	-3.6	27.1

are due to dispersion. In contrast, at $(k\tilde{r}X) = 6$ the system is robust and even though dispersion is considerably higher, there is good visual agreement between IWB and the reference, as expected.

D. Reproduction Error

To quantify the effect of dispersion on the final sound pressure at the ear, a measure of total reproduction error can be formulated as follows:

$$\epsilon_d(k) = \frac{\|\mathbf{p}^l - \tilde{\mathbf{p}}^l\|_2}{\|\mathbf{p}^l\|_2}, \quad (41)$$

where $\|\cdot\|_2$ denotes the 2-norm and

$$\mathbf{p}^l = [p^l(k, \Omega_1), p^l(k, \Omega_2), \dots, p^l(k, \Omega_S)]^T \quad (42)$$

denotes a frequency dependent vector of left-ear NHA predictions for S incidence directions at wavenumber k . In a similar fashion, $\tilde{\mathbf{p}}^l$ is constructed using NHA with a numerical wavenumber, \tilde{k} . Table I shows the overall reproduction error, $\epsilon_d(k)$, in decibels, for a number of characteristic frequencies and schemes. For the non-regularized cases, the error decreases up to $(k\tilde{r}X) = N$, due to the low robustness of the array¹. Above $(k\tilde{r}X) = N$, the error rises again because of numerical dispersion, which begins to play a prominent role at high frequencies. The contribution of spatial aliasing must be marginal compared to dispersion (see Fig. 3), as otherwise one would not expect such high variance in the reproduction error between the FDTD schemes at $(k\tilde{r}X) > N$.

For the regularized cases, the reproduction error becomes unacceptable only above $(k\tilde{r}X) > 2N$, which corresponds to $f/f_s = 0.38$ in Fig. 5. This suggests that with the aid of regularization, bandwidth constraints are largely imposed by the existence of numerical dispersion, and not by the array structure or processing algorithms.

VI. SIMULATION STUDY

So far, results have been presented for situations in which the sound field was assumed to be composed of a single plane wave. While these situations are useful for systematically studying the performance of the proposed array, they are not representative of realistic acoustic scenarios, most of which involve a continuum of plane waves. In a dispersive medium,

¹This is equivalent to $f/f_s = 0.19$, 0.11 and 0.15 for the IWB, SRL and IISO schemes, respectively, since at $(k\tilde{r}X) = N$ the frequency is $f/f_s = N\lambda/(2\pi\tilde{r})$.

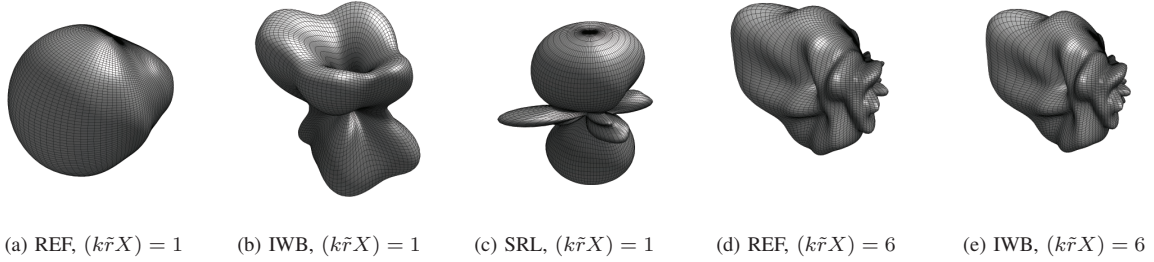


Fig. 6. Numerical HRTF analysis of 338 incidence directions at different frequencies and finite difference schemes. (a) the reference condition (REF) that has an ideal propagation wavenumber for $(k\tilde{r}X) = 1$, (b) the interpolated wideband scheme (IWB) for $(k\tilde{r}X) = 1$, (c) the standard rectilinear scheme (SRL) for $(k\tilde{r}X) = 1$, (d) the reference condition (REF) for $(k\tilde{r}X) = 6$, and (e) the interpolated wideband scheme (IWB) for $(k\tilde{r}X) = 6$. All volumetric arrays have a radius of $r\tilde{r} = 10$ nodes and are decomposed at $N = 12$.

each of these waves will propagate in accordance with their direction- and frequency- dependent phase velocities and will, therefore, have a different contribution to the overall input error at the array. Thus, for completeness, it is useful to perform a simulation study for a sound field that is composed of a number of reflections.

A domain of $3 \times 3 \times 3$ m was discretized in a grid resolution of $X = 10$ mm, corresponding to a total of 27×10^6 nodes. A sound source was positioned at a radial distance of 150 nodes (1.5 m) from a FVS array of $\tilde{r} = 10$ nodes. An IWB finite difference scheme was employed, in which the walls were modeled using frequency-independent boundary conditions, in accordance with [1], with a boundary impedance matching a wall absorption coefficient of $\alpha = 0.1$. The FDTD model was excited using a physically-constrained source based on an impulse response of a 32^{nd} order maximally-flat lowpass filter [33], with a cutoff frequency of $f/f_s = 0.186$ (corresponding to 6300 Hz) at the 2% dispersion error limit for the IWB scheme [1].

Additionally, the plane wave density function of a similar simulation setup was computed using the image source method (ISM) [34], [35]. For a simple rectilinear room with reflective boundaries, the ISM approaches an exact solution to the wave equation, and as such, serves as an ideal basis for comparison. Further, as directional information is inherently available in the ISM, it is straightforward to compute a plane wave density function without the need to explicitly simulate an array in the room [35], [36]. This avoids any errors related to array processing, and is the main reason for choosing the ISM here as a reference model. Since the array processing method described in this paper does not modify the core FDTD algorithm, its applicability to more complex acoustic scenarios can be directly inferred from the validation results.

Room impulse responses from both simulations were computed for the direct wave and for first order reflections of the sound field. In the case of the FDTD simulation, this was accomplished by choosing a simulation length such that only the direct wave and first-order reflections are included in the obtained data. Figure 7 shows the results of a PWD for FDTD and ISM sound fields at $k\tilde{r}X = N = 12$. In both cases, the directions of the direct component and the six first-order reflections of the sound field are clearly visible. Some

additional scattered energy can be seen in the PWD of the FDTD model. This is attributed to the early part of second-order reflections, which could not be entirely windowed-out from the modeled FDTD signals, but are otherwise not computed by the ISM model.

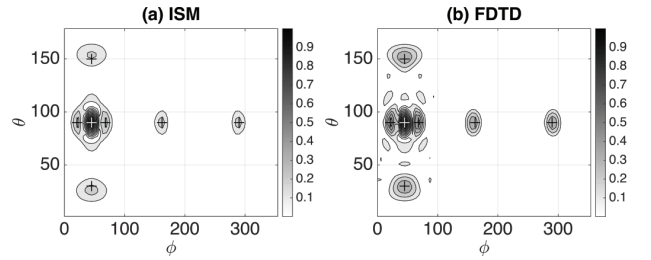


Fig. 7. Magnitude of the plane wave density function, $a(k, \Omega)$, at $k\tilde{r}X = N$, shown for a rectangular room solved with (a) the image source method (ISM) and (b) the FDTD method. The true direction of the modeled reflections are marked with a "+" symbol. For readability, the angle axes are shown in degrees.

Next, a Binaural Room Impulse Response (BRIR) was computed from the resulting plane wave densities, by employing (37). The ISM-based BRIR was additionally convolved with the FDTD excitation signal, to enable a clear comparison between the two methods. Figure 8 shows the reference ISM BRIR (solid line) against the modeled FDTD BRIR (dashed line) for incidence angles of $\Omega_0 = (\pi/2, 0)$ (upper pane) and $\Omega_0 = (\pi/2, \pi/4)$ (lower pane). For axial incidence, the two curves are in complete agreement, whereas for $\phi = \pi/4$ some small discrepancies are evident. These discrepancies may be attributed to the angle-dependency of the deviation between the Discrete Green's Function [37], which the source signal is effectively convolved with for the FDTD result, and its theoretical counterpart, with which it is convolved when employing the ISM. For the direct component of the sound field, this was verified by visually inspecting the raw (unprocessed) signals at the receivers, which also did not feature the small ripple evident in the ISM results. This phenomenon can largely be considered as a source excitation issue and, as such, is not a measure of the array processing. Other deviations could be related to sub-optimal processing of the array input data at $(k\tilde{r}X) \ll N$ (see Fig. 4c), which effectively impacts the directivity pattern of the underlying PWD beamformer.

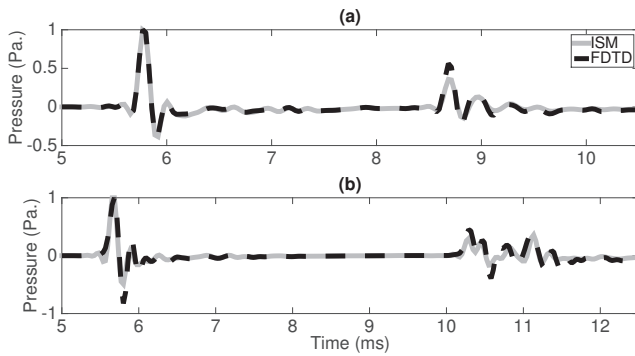


Fig. 8. Left-ear Binaural Room Impulse Responses (BRIRs) for the ISM simulation (solid line) and FDTD simulation (dashed line), shown for (a) $\Omega_0 = (\pi/2, 0)$ and (b) $\Omega_0 = (\pi/2, \pi/4)$. Responses are normalized for visual clarity. Note that the peak of the pulses are delayed compared to an ideal arrival time corresponding to a distance of 1.5m. This is due to the shape of the excitation signal itself, in which the peak of the pulse is time-shifted due to causal FIR filtering [33], and to additional delays inherent in the HRTFs with which the signals are filtered. To enable a clear visual comparison, results of the ISM were convolved with the same excitation signal used in the FDTD simulation.

VII. CONCLUSION

In this paper, a method for rendering a binaural response from an FDTD simulation was presented. A volumetric spherical array was employed for performing a PWD, followed by binaural synthesis in the SH domain. While the spherical array paradigm has been shown to be a suitable choice for performing a PWD, obtaining plane-wave components from an FDTD grid could also be accomplished by other means, such as virtual speaker arrays [14], convolution with directional derivatives [38], or by employing the Kirchhoff-Helmholtz integral equation [39]. Many of such methods would benefit from higher-order spatial interpolation on the pressure grid, but the literature is unclear yet as to how this would help to address the issue of proper regularization in the subsequent processing, which invariably is one of the key challenges in performing a broadband PWD. As such, a comprehensive study of different methods for applying a PWD to an FDTD grid remains an interesting topic for future research.

The volumetric array proposed in this paper can be seen as a special case of a spherical shell array [19], in which the entire array volume is utilized. Tools for analyzing such arrays in terms of robustness, aliasing and numerical dispersion were introduced and validated. In the current study, the SH decomposition order was limited to $N = 12$, although, in theory, it is only constrained by spatial aliasing and the desired frequency bandwidth. In practice, this bandwidth will also be determined by the amount of dispersion that is numerically or perceptually acceptable.

The method proposed in this paper operates in the SH domain. It can therefore potentially pave the way to modeling other types of receivers, e.g. directional microphones [38], through spherical beamforming techniques. This also has the added benefit of making sound field transformations more convenient. For example, to render signals for motion-tracked binaural auralizations, it is required to compute the binaural response at a large number of head rotations. When the plane-

wave density function is known, this can be simply achieved by multiplying with a Wigner-D function [40] in the spherical harmonics domain.

For loudspeaker-based auralization, the plane wave density function can be used to render signals for sound reproduction using higher order ambisonics and wavefield synthesis. Conversion of the plane-wave density function into 3D ambisonic signals only requires a translation of nomenclature. This can be accomplished simply by multiplication of a_{nm} with a transformation matrix, converting coefficients of the complex spherical harmonics into those of real spherical harmonics, as typically used in the ambisonics literature. If a 2D representation is desired, then conversion into cylindrical harmonics is required (see, for example [14]). Obtaining signals for wavefield synthesis can also be accomplished through Rayleigh's first integral. However, this may require a projection of the sound field onto the horizontal plane, as formulated in [41].

Examples of binaural auralizations generated using the method presented in this paper are available online: <http://www.ee.bgu.ac.il/~sheaffer/binaural.html>

ACKNOWLEDGMENT

The authors would like to thank David Alon for the insightful discussions on broadband array processing and Hai Morgenstein for commenting on the manuscript. The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 609465 as part of the Embodied Audition for RobotS (EARS) project.

REFERENCES

- [1] K. Kowalczyk and M. van Walstijn, "Room acoustics simulation using 3-D compact explicit FDTD schemes," *IEEE Trans. Audio Speech and Lang. Proc.*, vol. 19, no. 1, pp. 34–46, 2011.
- [2] T. Van Renterghem, E. Salomons, and D. Botteldooren, "Parameter study of sound propagation between city canyons with a coupled FDTD-PE model," *Applied Acoustics*, vol. 67, pp. 487–510, 2006.
- [3] S. Bilbao, *Numerical Sound Synthesis: Finite Difference Schemes and Simulation in Musical Acoustics*. Wiley, 2009.
- [4] M. Kleiner, B.-I. Dalenbäck, and P. Svensson, "Auralization—an overview," *J. Audio Eng. Soc.*, vol. 41, no. 11, pp. 861–875, 1993.
- [5] D. T. Murphy and M. Beeson, "The KW-boundary hybrid digital waveguide mesh for room acoustics applications," *IEEE Trans. Audio Speech and Lang. Proc.*, vol. 15, no. 2, pp. 552–564, 2007.
- [6] C. J. Webb and S. Bilbao, "Binaural simulations using audio rate FDTD schemes and CUDA," in *Proc. of the 15th Int. Conference on Digital Audio Effects (DAFx-12)*, York, United Kingdom, 2012.
- [7] J. Sheaffer, C. Webb, and B. M. Fazenda, "Modelling binaural receivers in finite difference simulation of room acoustics," in *Proceedings of Meetings on Acoustics*, vol. 19, no. 1. Acoustical Society of America, 2013, p. 015098.
- [8] T. Xiao and Q. H. Liu, "Finite difference computation of head-related transfer function for human hearing," *The Journal of the Acoustical Society of America*, vol. 113, no. 5, pp. 2434–2441, 2003.
- [9] P. Mokhtari, H. Takemoto, R. Nishimura, and H. Kato, "Computer simulation of HRTFs for personalization of 3D audio," in *Universal Communication, 2008. ISUC'08. Second International Symposium on*. IEEE, 2008, pp. 435–440.
- [10] A. Meshram, R. Mehra, and D. Manocha, "Efficient HRTF computation using adaptive rectangular decomposition," in *AES 55th International Conference, Helsinki, Finland*, 2014.
- [11] S. Bilbao, "Modeling of complex geometries and boundary conditions in finite difference/finite volume time domain room acoustics simulation," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 21, no. 7, pp. 1524–1533, 2013.

- [12] V. R. Algazi, R. O. Duda, and D. M. Thompson, "Motion-tracked binaural sound," *Journal of the Audio Engineering Society*, vol. 52, no. 11, pp. 1142–1156, 2004.
- [13] M. A. Poletti and U. P. Svensson, "Beamforming synthesis of binaural responses from computer simulations of acoustic spaces," *The Journal of the Acoustical Society of America*, vol. 124, no. 1, pp. 301–315, 2008.
- [14] B. Støfringsdal and P. Svensson, "Conversion of discretely sampled sound field data to auralization formats," *J. Audio Eng. Soc.*, vol. 54, no. 5, pp. 380–400, 2006.
- [15] J. C. Strikwerda, *Finite difference schemes and partial differential equations*. Siam, 2004.
- [16] L. Savioja, D. Manocha, and M. Lin, "Use of gpus in room acoustic modeling and auralization," in *Proc. Int. Symposium on Room Acoustics*, 2010.
- [17] T. Yokota, S. Sakamoto, and H. Tachibana, "Sound field simulation method by combining finite difference time domain calculation and multi-channel reproduction technique," *Acoustical Science and Technology*, vol. 25, no. 1, pp. 15–23, 2004.
- [18] A. Southern, D. T. Murphy, and L. Savioja, "Spatial encoding of finite difference time domain acoustic models for auralization," *IEEE Trans. Audio Speech and Lang. Proc.*, vol. 20, no. 9, pp. 2420–2432, 2012.
- [19] B. Rafaely, "The spherical-shell microphone array," *IEEE Trans. Audio Speech and Lang. Proc.*, vol. 16, no. 4, pp. 740–747, 2008.
- [20] J. Sheaffer, M. van Walstijn, B. Rafaely, and K. Kowalczyk, "A spherical array approach for simulation of binaural impulse responses using the finite difference time domain method," in *Proceedings of Forum Acusticum*, 2014.
- [21] E. G. Williams, *Fourier acoustics: sound radiation and nearfield acoustical holography*. academic press, 1999.
- [22] B. Rafaely, "Plane-wave decomposition of the sound field on a sphere by spherical convolution," *J. Acoust. Soc. Am.*, vol. 116, no. 4, pp. 2149–2157, 2004.
- [23] B. Rafaely and A. Avni, "Interaural cross correlation in a sound field represented by spherical harmonics," *J. Acoust. Soc. Am.*, vol. 127, no. 2, pp. 823–828, 2010.
- [24] B. Bernschütz, C. Pörschmann, S. Spors, S. Weinzierl, and B. der Verstärkung, "Softlimiting der modalen amplitudenverstärkung bei sphärischen mikrofonarrays im plane wave decomposition verfahren," *Proceedings of the 37. Deutsche Jahrestagung für Akustik (DAGA 2011)*, pp. 661–662, 2011.
- [25] T. Rettberg and S. Spors, "Time-domain behaviour of spherical microphone arrays at high orders."
- [26] B. Rafaely, "Analysis and design of spherical microphone arrays," *IEEE Trans. on Speech and Audio Proc.*, vol. 13, no. 1, pp. 135–143, 2005.
- [27] B. Rafaely, B. Weiss, and E. Bachmat, "Spatial aliasing in spherical microphone arrays," *IEEE Transactions on Signal Processing*, vol. 55, no. 3, pp. 1003–1010, 2007.
- [28] D. L. Alon and B. Rafaely, "Spherical microphone array with optimal aliasing cancellation," in *2012 IEEE 27th Convention of Electrical & Electronics Engineers in Israel (IEEEI)*. IEEE, 2012, pp. 1–5.
- [29] I. Balmages and B. Rafaely, "Open-sphere designs for spherical microphone arrays," *IEEE Trans. Audio Speech and Lang. Proc.*, vol. 15, no. 2, pp. 727–732, 2007.
- [30] L. Savioja and V. Valimaki, "Interpolated rectangular 3-D digital waveguide mesh algorithms with frequency warping," *IEEE Trans. Audio Speech and Lang. Proc.*, vol. 11, no. 6, pp. 783–790, 2003.
- [31] A. Southern, D. Murphy, T. Lokki, and L. Savioja, "The perceptual effects of dispersion error on room acoustic model auralization," in *Proc. Forum Acusticum, Aalborg, Denmark*, 2011, pp. 1553–1558.
- [32] F. L. Wightman and D. J. Kistler, "Monaural sound localization revisited," *The Journal of the Acoustical Society of America*, vol. 101, no. 2, pp. 1050–1063, 1997.
- [33] J. Sheaffer, M. van Walstijn, and B. Fazenda, "Physical and numerical constraints in source modeling for finite difference simulation of room acoustics," *J. Acoust. Soc. Am.*, vol. 135, no. 1, pp. 251–261, 2014.
- [34] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, 1979.
- [35] A. Wabnitz, N. Epain, C. Jin, and A. van Schaik, "Room acoustics simulation for multichannel microphone arrays," in *Proceedings of the International Symposium on Room Acoustics*, 2010.
- [36] D. P. Jarrett, E. A. Habets, M. R. Thomas, and P. A. Naylor, "Simulating room impulse responses for spherical microphone arrays," in *ICASSP*, 2011, pp. 129–132.
- [37] R. Kastner, "A multidimensional z-transform evaluation of the discrete finite difference time domain green's function," *Antennas and Propagation, IEEE Transactions on*, vol. 54, no. 4, pp. 1215–1222, 2006.
- [38] H. Hacıhabiboglu, B. Gunel, and Z. Cvetkovic, "Simulation of directional microphones in digital waveguide mesh-based models of room acoustics," *IEEE Trans. Audio Speech and Lang. Proc.*, vol. 18, no. 2, pp. 213–223, 2010.
- [39] S. Bleda, J. Escolano, J. J. Lopez, and B. Pueo, "An approach to discrete-time modelling auralization for wave field synthesis applications," in *Audio Engineering Society Convention 118*. Audio Engineering Society, 2005.
- [40] B. Rafaely and M. Kleider, "Spherical microphone array beam steering using Wigner-D weighting," *IEEE Signal Processing Letters*, vol. 15, pp. 417–420, 2008.
- [41] J. Ahrens and S. Spors, "Wave field synthesis of a sound field described by spherical harmonics expansion coefficients," *J. Acoust. Soc. Am.*, vol. 131, no. 3, pp. 2190–2199, 2012.