Source and Listener Directivity for Interactive Wave-based Sound Propagation

Ravish Mehra, Lakulish Antani, Sujeong Kim, and Dinesh Manocha

Fig. 1. Our source and listener directivity is modifiable at runtime and can be analytical, data-driven, rotating, or can be any time-varying function. We demonstrate realistic acoustic effects produced by wave-based sound propagation for directional sources and listeners in the following scenarios: (left) amphitheater, (center) living room and (right) reservoir (Half-life 2).

Abstract—We present an approach to model dynamic, data-driven source and listener directivity for interactive wave-based sound propagation in virtual environments and computer games. Our directional source representation is expressed as a linear combination of elementary spherical harmonic (SH) sources. In the preprocessing stage, we precompute and encode the propagated sound fields due to each SH source. At runtime, we perform the SH decomposition of the varying source directivity interactively and compute the total sound field at the listener position as a weighted sum of precomputed SH sound fields. We propose a novel plane-wave decomposition approach based on higher-order derivatives of the sound field that enables dynamic HRTF-based listener directivity at runtime. We provide a generic framework to incorporate our source and listener directivity in any offline or online frequency-domain wave-based sound propagation algorithm. We have integrated our sound propagation system in Valve's Source game engine and use it to demonstrate realistic acoustic effects such as sound amplification, diffraction low-passing, scattering, localization, externalization, and spatial sound, generated by wave-based propagation of directional sources and listener in complex scenarios. We also present results from our preliminary user study.

Index Terms—Sound propagation, directivity, spatial sound, plane-wave decomposition, helmholtz equation

1 Introduction

Realistic sound effects are extremely critical in virtual reality simulations to improve the sense of presence and realism of virtual environments [3, 31, 13]. They augment the visual sense of the user increasing situational awareness and improves the ability to localize sound [4]. Studies in audio-visual cross-modal perception have shown that high quality sound rendering can increase the quality perception of visual rendering [31]. For generating realistic sound effects, a key issue is performing sound propagation for directional sources and listeners. Most sound sources we come across in real life, ranging from human voices through speaker systems, machine noises, and musical instruments, are directional sources that have a specific *directivity pattern* [11, 18]. Source directivity has a significant impact on the propagation of sound and the corresponding acoustics of the environments [33] that is noticeable in everyday life: a person talking towards/away from a listener, the positioning of different types of musical instruments in an orchestra [18], and good-sounding places (sweet spots) in front of the television in the living room. Analogous to directional sources, listeners also do not receive sound equally from all directions. The human auditory system obtains significant directional cues from the subtle differences in the sound received at the left and right ear, caused by the scattering of sound around the head. This listener directivity is encoded as the head-related transfer function (HRTF). Spatial sound based on listener directivity can be used to enhance a user's immersion in the virtual environment by providing binaural cues corresponding to the direction the sound is coming from, thereby enriching the experience [3].

Existing sound propagation techniques, broadly classified into geometric and wave-based techniques, for handling dynamic source and listener directivity have many limitations. Geometric techniques can easily handle high-frequency source and listener directivity for offline and interactive applications [6]. However, due to the inherent assumption of rectilinear propagation i.e. sound waves travel as rays, in geometric techniques, the modeling wave effects, such as diffraction and interference, at low frequencies remains a significant challenge. This becomes a limiting factor for low-frequency directional sources (e.g. human voices) and low-frequency listener directivity effects (e.g. diffraction around the head). Wave-based techniques can accurately perform sound propagation at low frequencies, but their computational complexity increases significantly for high frequencies. Current interactive wave-based techniques [10, 28, 16] have a high precomputation overhead and can only model source directivity during the offline computation stage. As a result, the source directivity gets baked (precomputed and stored) into the final solution, and it is not possible to modify the directivity pattern at runtime for interactive applications e.g. a rotating siren or a person covering his/her mouth. Additionally, integrating listener directivity into wave-based techniques requires a plane-wave decomposition of the sound field [36]. Previous techniques for performing plane-wave decomposition [23, 27, 36] are computationally expensive and limited to offline applications.

In this paper, we address the problem of incorporating dynamic

- *Ravish Mehra is with UNC Chapel Hill. E-mail: ravishm@cs.unc.edu.*
- *Lakulish Antani is with UNC Chapel Hill. E-mail: lakulish@cs.unc.edu.*
- *Sujeong Kim is with UNC Chapel Hill. E-mail: sujeong@cs.unc.edu.*
- *Dinesh Manocha is with UNC Chapel Hill. E-mail: dm@cs.unc.edu.*

Manuscript received 12 September 2013; accepted 10 January 2014; posted online 29 March 2014; mailed on 1 May 2014.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

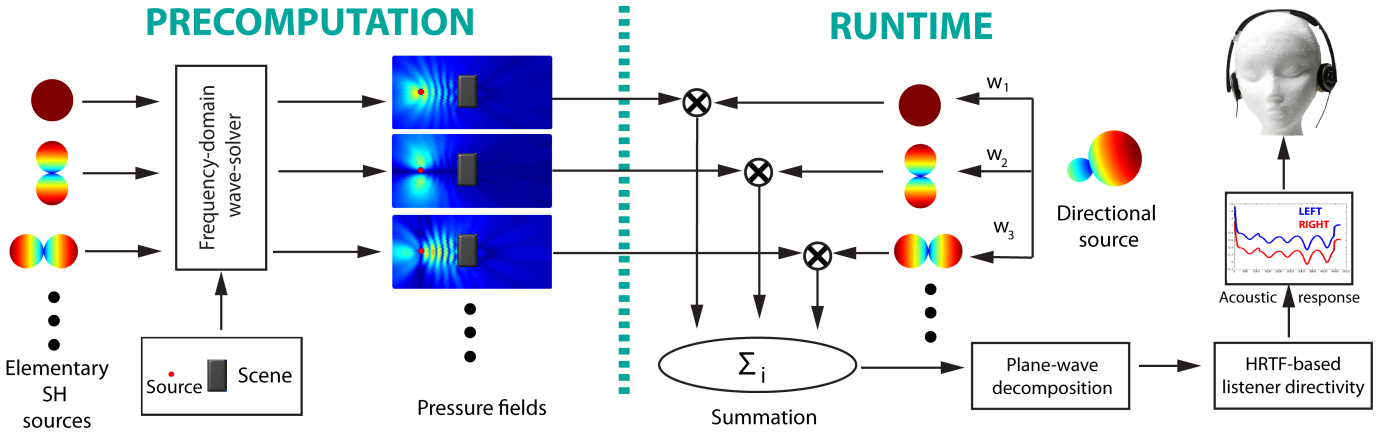


Fig. 2. **Overview of our approach.**

source and listener directivity in interactive wave-based sound propagation techniques. Figure 2 gives an overview of our approach. Given a scene and a source position, we precompute a set of pressure fields due to elementary spherical harmonic (SH) sources using a frequency-domain wave-based sound propagation technique. Next, we encode these pressure fields in basis functions (e.g. multipoles) and store them for runtime use. Given the dynamic source directivity at runtime, we perform a SH decomposition of the directivity to compute the corresponding SH coefficients. The final pressure field is computed as a summation of the pressure fields due to SH sources evaluated at the listener position weighted by the appropriate SH coefficients. In order to incorporate dynamic listener directivity in wave-based techniques, we propose an interactive plane-wave decomposition approach based on derivatives of the pressure field. Acoustic responses for both ears are computed at runtime by using this efficient plane-wave decomposition of the pressure field and the HRTF-based listener directivity. These binaural acoustic responses are convolved with the (dry) audio to compute the propagated spatial sound at the listener position.

Main Results: Our main contributions are:

1. *Dynamic, data-driven source directivity* modifiable at runtime. We precompute and store the propagated sound fields due to SH sources and use them at runtime to compute the sound field due to a directional source.

2. *Efficient plane-wave decomposition* based on pressure derivatives that enables *dynamic HRTF-based listener directivity* at runtime. Our formulation is applicable to interactive applications and supports listener head rotation and allows the use of personalized HRTFs without recomputing the sound pressure fields.

3. *General framework* to integrate our source and listener directivities into *any* offline or interactive frequency-domain wave-based propagation algorithm.

4. *Real-time, memory-efficient* sound rendering system. We demonstrate realistic acoustic effects from directional sources and listener in complex scenarios.

We have demonstrated that our directivity approach works for both offline and interactive wave-based sound propagation techniques by incorporating it into the state-of-the-art offline boundary element method [14] and the interactive equivalent source technique [16]. We have also integrated our runtime system with Valve's Source™ game engine. We demonstrate acoustic effects from both source and listener directivity on a variety of scenarios, such as people talking on the street, loudspeakers between buildings, a television in a living room, a helicopter in a rocky outdoor terrain, a bell tower in a snow-covered town, a rotating siren, and musical instruments in an amphitheater. We also validate our results with the ground-truth responses computed analytically using the offline Biot-Tolstoy-Medwin technique [32]. Our method enables accurate wave-based sound propagation for dynamic source and listener directivities and can handle moving directional sources and a moving directional listener in interactive applications. We have also conducted a preliminary user study to evaluate the results generated by our source and listener directivity approach.

2 Prior Work

We briefly discuss the prior work in this area.

Sound propagation The techniques to perform sound propagation can be classified into two categories - *geometric acoustic* and *wave-based*. Geometric acoustic (GA) techniques are based on the assumption of rectilinear propagation of sound waves, which is valid only at high frequencies. Many GA techniques have been developed, mostly based on image sources, ray tracing, and beam tracing [34, 6]. Recent techniques include diffraction modeling using the Uniform Theory of Diffraction [6] and the Biot-Tolstoy-Medwin model [32]. Wave-based techniques numerically solve the acoustic wave equation [24], capturing all wave phenomena, such as diffraction, scattering, and interference. Wave-based techniques include the finite difference method, the finite element method, the boundary element method, the equivalent source method, and the spectral method [21].

Source directivity Meyer et al. [18] measure the directivity of brass, woodwind and string instruments in an anechoic chamber. Recently, directivities of male and female singing voices have also been measured [11]. These datasets are available online. Physically-based sound synthesis algorithms can simulate sound radiation from directional sources [5]. However, these techniques model only free-space sound radiation. In contrast, our goal is to model propagation effects (reflections, diffraction, reverberation) from other objects in the scene. Interactive GA techniques can incorporate high-frequency source directivities at runtime [6, 22]. Interactive wave-based sound propagation techniques [28, 16] can handle elementary directional sources such as monopoles, dipoles, quadrupoles, and their linear combinations. Other techniques have been proposed to incorporate measured directivities in wave-based techniques [8, 30].

Listener directivity A *head-related transfer function* (HRTF) describes the effect of the listener's outer ear, head and body on the sound arriving along different directions. Measurements to compute HRTF are performed in controlled environments and the recorded data is available online [1]. Interactive GA techniques can incorporate high-frequency HRTF-based listener directivity at runtime. However, integrating HRTFs into wave-based techniques involves computation of propagation directions using *plane wave decomposition*. Prior plane-wave decomposition techniques either use spherical convolution [23, 27] or solve a linear system [36], and are computationally expensive. Interactive wave-based techniques resort to simpler

listener directivity models based on a spherical head and a cardioid function [28, 16]. However, these simplified models are not accurate for sound localization and externalization, both of which are necessary for immersion in virtual environments [3].

Spherical harmonics (SH) in acoustics SH form an orthonormal basis for functions defined on a sphere. Warfusel et al. [35] generate source directivity from a combination of elementary SH directivities using a 3D loudspeaker array. However, the system does not support dynamic, data-driven source directivity at runtime, and only handles morphing between elementary directivities. Also, listener directivity was handled by recording binaural responses of the room at a fixed listener head position and orientation. Recording binaural responses for all possible listener positions and orientations becomes an infeasible 5D sampling problem, requiring huge memory and manual effort. Noisternig [19] proposed a source directivity approach for GA techniques based on discrete SH transform. However, geometric techniques cannot model directivity at low frequencies. In recent years, several SH-based models have been proposed for representing HRTFs [27, 25]. Our work is orthogonal to these techniques. To compute spatial sound, a dot product of SH coefficients of the HRTF and the plane wave decomposition of pressure field is required [27]. These techniques focus on the first part - computing better HRTF coefficients. Our main contribution lies in the second part - efficiently computing SH coefficients of the plane-wave decomposition at interactive rates.

3 Source Directivity

We discuss a far-field source representation that can be used to efficiently handle dynamic, data-driven source directivity. We then present an approach to incorporate our source representation into a general frequency-domain wave-based propagation technique. All the variables used henceforth, except the SH basis functions, positions and speed of sound, are frequency-dependent. For the sake of brevity these dependencies are not mentioned explicitly.

3.1 Source representation

The radiation pattern of a generic directional source can be expressed using the one-point multipole expansion [20] as:

$$s(\mathbf{x},\mathbf{y}) = \sum_{l=0}^{L-1}\sum_{m=-l}^{l} b_{lm}\, h_l^2(2\pi\nu r/c)\, Y_{lm}(\theta,\phi), \tag{1}$$

where $s(\mathbf{x},\mathbf{y})$ is the pressure field at point $\mathbf{x}$ of the directional source centered at point $\mathbf{y}$, $h_l^2(.)$ are the spherical Hankel functions of second kind, $Y_{lm}(.)$ are complex-valued SH basis functions, (r,θ,ϕ) is the vector $(\mathbf{x}-\mathbf{y})$ expressed in spherical coordinates, b_{lm} are weights and L is order of the expansion, ν is the frequency, and c is the speed of sound. This source formulation is valid in both near and far-field[1].

Our choice of the source representation for directional sources is motivated by the measured directivity data that is currently available for real-world sound sources. Most available measurements have been collected by placing sources in an anechoic chamber and recording their directivity by rotating microphones every few degrees at a fixed distance from the source. Typically, these measurements are carried out at a distance of a few meters, which corresponds to the far-field for the frequencies emitted by these sources[2]. Keeping this in mind, we chose a source representation that corresponds to the far-field radiation pattern of a directional source. Under far-field approximation, $h_l^2(z) \approx i^{l+1}e^{-iz}/z$ where $i=\sqrt{-1}$ simplifying equation 1 to following source representation [17]:

$$s(\mathbf{x},\mathbf{y}) = \frac{e^{-i2\pi\nu r/c}}{r}\sum_{l=0}^{L-1}\sum_{m=-l}^{l} a_{lm}Y_{lm}(\theta,\phi), \tag{2}$$

$$= \sum_{l=0}^{L-1}\sum_{m=-l}^{l} a_{lm}s_{lm}(\mathbf{x},\mathbf{y}) = \frac{e^{-i2\pi\nu r/c}}{r}D(\theta,\phi), \tag{3}$$

where $a_{lm} = b_{lm}i^{l+1}c/(2\pi\nu)$ are the SH coefficients, $s_{lm}(\mathbf{x},\mathbf{y}) = \frac{e^{-i2\pi\nu r/c}}{r}Y_{lm}$ are the elementary SH sources, and $D(\theta,\phi) = \sum\sum a_{lm}Y_{lm}(\theta,\phi)$ is the directivity function at frequency ν. This directivity function can be specified for each frequency either analytically or measured or simulated at discrete sample directions. Depending on the data, the directivity function can be complex-valued (both magnitude and phase) or real-valued (magnitude-only) function. Typically, the measured data is magnitude-only and available as directivities averaged over octave-wide frequency bands [26]. We compute the coefficients a_{lm} of the source representation from the directivity function $D(\theta,\phi)$ by standard SH projection methods [7].

3.2 Frequency-domain sound propagation

Sound wave propagation in the frequency domain can be expressed as a boundary value problem using the Helmholtz equation:

$$\nabla^2 p + \frac{\omega^2}{c^2}p = 0 \text{ in } \Omega, \tag{4}$$

where $p(\mathbf{x})$ is the complex-valued pressure field (sound field) at frequency ν, $\omega = 2\pi\nu$ is the angular frequency, Ω is the propagation domain, and ∇^2 is the Laplacian operator. This equation can be solved using any frequency-domain, wave-based propagation technique, including the boundary element method, the finite element method or the equivalent source method. We describe our approach for incorporating the directional source representation in a general frequency-domain, wave-based sound propagation technique. The steps outlined below are repeated for frequency samples in the range $[0,\nu_{\max}]$, where $\nu_{\max}$ is the maximum frequency simulated.

The linearity of the Helmholtz equation implies that the pressure field of a linear combination of sources is a linear combination of their respective pressure fields [24]. Our source representation for directional sources is a linear combination of elementary SH sources $s_{lm}(\mathbf{x},\mathbf{y})$ with different weights a_{lm} (equation 3). Therefore, for a given scene, if we compute the pressure field $p_{lm}(\mathbf{x})$ corresponding to each of the elementary SH sources $s_{lm}(\mathbf{x},\mathbf{y})$, then the pressure field $p(\mathbf{x})$ due to any directional source $s(\mathbf{x},\mathbf{y})$ can be expressed as the linear combination of the precomputed pressure fields of the elementary SH sources $p_{lm}(\mathbf{x})$ with the same weights a_{lm}:

$$\underbrace{\sum_l\sum_m a_{lm}s_{lm}(\mathbf{x},\mathbf{y})}_{s(\mathbf{x},\mathbf{y})} \longrightarrow \underbrace{\sum_l\sum_m a_{lm}p_{lm}(\mathbf{x})}_{p(\mathbf{x})}.$$

The pressure fields for elementary SH sources can be computed using any wave-based sound propagation technique. In the case of interactive applications, this computation is performed during the preprocessing stage, and the resulting pressure field data is efficiently encoded and stored. This pressure field data completely defines the acoustic response to any directional source at the given position up to the SH approximation order. At runtime, the specified source directivity $D(\theta,\phi)$ is decomposed into a SH-based representation, and the resulting weights a_{lm} are used to compute the final acoustic response at listener position, as described above. Details for handling rotating and dynamic directivity are discussed later in Section 6.

4 Listener Directivity

We now describe the method for computing spatial sound using efficient plane-wave decomposition of the pressure field and HRTF-based listener directivity.

[1] Far-field refers to the region of space where the distance d of any point in that region to the source is greater than the wavelength λ of the sound emitted by the source. The complementary region is the near-field [12, p. 165].

[2] For a distance $d > 3.43$m, corresponds to far-field for all the frequencies $\nu > 100$ Hz (for all wavelengths $\lambda = c/\nu < 3.43$m).

4.1 Spatial sound

In the frequency domain, the global sound field can be expressed as a superposition of pressure due to plane waves [36]:

$$p(\mathbf{x}) = \frac{1}{4\pi}\int_S \psi_{\mathbf{s}}(\mathbf{x})\mu(\mathbf{s})d\mathbf{s}, \tag{5}$$

where $\psi_{\mathbf{s}}(\mathbf{x}) = e^{ik\mathbf{s}\cdot(\mathbf{x}-\mathbf{x}_0)}$ is the plane wave basis (also called *Herglotz wave basis*), $\mathbf{x}_0$ is the listener position, $\mathbf{s} = (s_x, s_y, s_z)$ is the unit vector in the direction of plane wave propagation, $k = 2\pi\nu/c$ is the wave number, and $\mu(\mathbf{s})$ is the *signature function* that specifies the complex-valued amplitude of the plane wave traveling along direction $\mathbf{s}$ for a given frequency ν. The pressure received at the left ear due to a plane wave traveling along direction $\mathbf{s}$ is given by: $H_L(\mathbf{s})\psi_{\mathbf{s}}(\mathbf{x}_0)$ where $H_L(\mathbf{s})$ is the HRTF function for the left ear. The total pressure received at the left ear is obtained by integration:

$$p_L = \frac{1}{4\pi}\int_S H_L(\mathbf{s})\psi_{\mathbf{s}}(\mathbf{x}_0)\mu(\mathbf{s})d\mathbf{s} = \frac{1}{4\pi}\int_S H_L(\mathbf{s})\mu(\mathbf{s})d\mathbf{s}, \tag{6}$$

since $\psi_{\mathbf{s}}(\mathbf{x}_0) = 1$. As discussed in [27], by using the SH decomposition of the HRTF $H_L(\mathbf{s}) = \sum_{l'}\sum_{m'}\ \beta^L_{l'm'}Y_{l'm'}(\mathbf{s})$ and the signature function $\mu(\mathbf{s}) = \sum_l\sum_m \alpha_{lm}Y^*_{lm}(\mathbf{s})$ (where Y^*_{lm} is conjugate SH), the above integral gets simplified to a dot product using the orthonormality property as $p_L = \frac{1}{4\pi}\sum_l\sum_m\ \beta^L_{lm}\alpha_{lm}$. Analogous equation can be derived for the right ear. Thus, spatial sound at both ears can be computed as a dot product of SH coefficients of the plane-wave decomposition and the HRTFs.

4.2 Plane-wave decomposition using derivatives

To support a moving listener with a rotating head (dynamic listener directivity), one needs to update the SH coefficients of the HRTFs β_{lm} and plane-wave decomposition α_{lm} interactively. Listener head rotation can be incorporated by applying SH rotation techniques to the SH coefficients of the HRTF. A dynamic HRTF can also be handled using fast SH decomposition techniques in the same manner as dynamic source directivity discussed before. However, the SH coefficients of the plane-wave decomposition will have to be recomputed as the listener moves at runtime. Previous techniques [23, 27, 36] cannot compute the plane-wave decomposition at interactive rates. We propose a novel method to compute the SH coefficients of the plane-wave decomposition at interactive rates using the derivatives of the pressure field.

Theorem : *The SH coefficients of the plane-wave decomposition of the pressure field can be expressed as a linear combination of pressure field derivatives at a single point.*

Proof : Given the polynomial expression of the n^{th} order and q^{th} degree SH as $Y_{nq}(\mathbf{s}) = A\sum_{(a,b,c)}\Gamma_{a,b,c}\ \ s_x^a\, s_y^b\, s_z^c$, where $\Gamma_{a,b,c}$ is a constant, $a \geq 0, b \geq 0, c \geq 0$, and $a+b+c = n$. For computing the SH coefficient of the plane-wave decomposition α_{nq}, we multiply both sides by $\alpha_{lm}Y^*_{lm}$ (where Y^*_{lm} is conjugate SH) and apply SH summation and integral operators to give

$$\sum_l\sum_m \alpha_{lm}\int Y_{nq}\,Y^*_{lm}d\mathbf{s} = A\sum_{(a,b,c)}\Gamma_{a,b,c}\ \ \sum_l\sum_m\alpha_{lm}\int s_x^a\, s_y^b\, s_z^c\, Y^*_{lm}d\mathbf{s}.$$

Using the orthonormality property of SH, we get

$$\alpha_{nq} = A\sum_{(a,b,c)}\Gamma_{a,b,c}\ \ \sum_l\sum_m\alpha_{lm}\int s_x^a\, s_y^b\, s_z^c\, Y^*_{lm}d\mathbf{s}. \tag{7}$$

We denote the partial a^{th} x-derivative, partial b^{th} y-derivative, partial c^{th} z-derivative of pressure field as $p^{(a,b,c)} = \frac{\partial^{a+b+c}p}{\partial x^a\partial y^b\partial z^c}$. The total pressure at point $\mathbf{x}$ (equation 5) along with the SH-expansion for signature function $\mu(s)$ gives

$$p(\mathbf{x}) = \frac{1}{4\pi}\sum_l\sum_m\alpha_{lm}\int_S \psi_{\mathbf{s}}(\mathbf{x})Y^*_{lm}(\mathbf{s})d\mathbf{s}.$$

On differentiating and evaluating the above expression at $\mathbf{x}_0$, we get

$$\frac{4\pi}{(ik)^{a+b+c}}\ \ p^{(a,b,c)}(\mathbf{x}_0) = \sum_l\sum_m\alpha_{lm}\int s_x^a\, s_y^b\, s_z^c\, Y^*_{lm}d\mathbf{s}.$$

Substituting above expression in equation (7), we get

$$\alpha_{nq} = A\sum_{(a,b,c)}\Gamma_{a,b,c}\ \ \frac{4\pi}{(ik)^{a+b+c}}\ \ p^{(a,b,c)}(\mathbf{x}_0). \tag{8}$$

The above expression relates the pressure field derivatives to the SH coefficients of the plane-wave decomposition at the listener position $\mathbf{x}_0$. This gives us an efficient method to compute the plane-wave decomposition using derivatives of the pressure field.

4.3 Frequency-domain sound propagation

At runtime, the pressure field derivatives at the listener position are computed interactively by differentiating the pressure basis functions of the wave-based propagation technique *analytically* rather than using a finite difference stencil (Section 5). Therefore, higher-order derivatives do not suffer from numerical instabilities allowing the SH coefficients of the plane-wave decomposition to be computed to higher-orders (equation 8). Finally, we compute the spatial sound for the left and right ears as a dot product of the SH coefficients of the plane-wave decomposition and the HRTFs.

Our approach provides the flexibility to use personalized (user-specific) HRTFs without recomputing the simulation results. In this case, only the SH coefficients of the HRTFs need to be updated. Our plane-wave decomposition method allows efficient computation of spatial sound, enabling the use of HRTF-based listener directivity for a moving listener at runtime. Head rotation is enabled by applying efficient SH rotation techniques [7] to the SH coefficients of the HRTFs. Therefore, our method can compute spatial sound for a head-tracked listener at interactive rates. This could result in better immersion in virtual- and augmented-reality applications and computer games.

5 Wave-based Sound Propagation

We discuss the integration of our source and listener directivity with frequency-domain, wave-based sound propagation techniques.

5.1 Boundary element method (BEM)

The boundary element method [14] is a numerical technique used for solving the 3D Helmholtz equation that accurately models sound propagation in indoor and outdoor spaces. BEM transforms the Helmholtz equation into the boundary integral equation, then solves for pressure and velocity on the boundary-and thereby pressure at any point in the domain.

Simulation Given an indoor or outdoor scene with source position $\mathbf{y}$, we solve the Helmholtz equation corresponding to the elementary SH source $s_{lm}(\mathbf{x},\mathbf{y})$ using BEM. This gives us pressure $q_{lm}(\mathbf{z})$ and velocity $v_{lm}(\mathbf{z})$ on the domain boundary ($\mathbf{z}$ is a point on the boundary). This step is repeated for all the elementary sources $s_{lm}(\mathbf{x},\mathbf{y})$ in the SH expansion (Equation 3). The pressure and velocity on the domain boundary is stored and then used at runtime for evaluating pressure and its high-order derivatives at the listener position.

Pressure evaluation We compute the pressure field $p_{lm}(\mathbf{x})$ due to the elementary SH sources $s_{lm}(\mathbf{x})$ by using the boundary integral equation:

$$p_{lm}(\mathbf{x}) = \int_S [G(\mathbf{x},\mathbf{z},\omega)q_{lm}(\mathbf{z}) - F(\mathbf{x},\mathbf{z})v_{lm}(\mathbf{z})]\,dS(\mathbf{z}) + s_{lm}(\mathbf{x},\mathbf{y}),$$

where $G(\mathbf{x},\mathbf{z},\omega) = \frac{1}{4\pi d}e^{i\omega d/c}$ is the Green's function ($d = \|x-z\|$), $F(\mathbf{x},\mathbf{z}) = \frac{\partial G(\mathbf{x},\mathbf{z},\omega)}{\partial n(\mathbf{z})}$ and $n(\mathbf{z})$ is normal at a boundary point $\mathbf{z}$. The final pressure field at the listener position $\mathbf{x}_0$ due to the directional source $s(\mathbf{x},\mathbf{y})$ is computed as a linear combination of the pressure fields of the elementary SH sources as $p(\mathbf{x}_0) = \sum_l\sum_m a_{lm}p_{lm}(\mathbf{x}_0)$.

Pressure derivative evaluation In order to produce spatial sound, we need to determine the coefficients of the plane wave decomposition using pressure derivatives (Equation 8). The first-order derivative of the pressure field due to elementary source $\partial p_{lm}(\mathbf{x})/\partial x$ is computed by differentiating the functions involved analytically:

$$\frac{\partial p_{lm}(\mathbf{x})}{\partial x} = \int_S \left[\frac{\partial G(\mathbf{x},\mathbf{z},\omega)}{\partial x} q_{lm}(\mathbf{z}) - \frac{\partial F(\mathbf{x},\mathbf{z})}{\partial x} v_{lm}(\mathbf{z}) \right] dS(\mathbf{z}) + \frac{\partial s_{lm}(\mathbf{x})}{\partial x}.$$

The first-order derivative of the total pressure field at the listener position $\mathbf{x}_0$ is computed as $\frac{\partial p(\mathbf{x}_0)}{\partial x} = \sum_l \sum_m a_{lm} \frac{\partial p_{lm}(\mathbf{x}_0)}{\partial x}$. Higher-order derivatives of the pressure field at the listener position can be computed in a similar manner.

5.2 Equivalent source method (ESM)

Equivalent source technique [16] can perform interactive wave-based sound propagation in large outdoor scenes. This technique decomposes the global pressure field of a scene into local per-object sound fields and inter-object interactions, which are precomputed offline. The pressure field is computed by solving a global linear system consisting of per-object and inter-object interactions; it is then encoded efficiently as the strengths of equivalent sources. At runtime, the acoustic response at a moving listener is efficiently computed by performing a fast summation over all the equivalent sources.

Preprocessing Assume a scene composed of κ objects, A_1, A_2, ..., A_κ and source position $\mathbf{y}$. During the preprocessing stage, we express the elementary SH source $s_{lm}(\mathbf{x},\mathbf{y})$ in terms of the incoming equivalent sources of these objects. Next, we perform the global solve step to compute the strengths of outgoing equivalent sources for all the objects $(C_{lm})_{A_1}$, $(C_{lm})_{A_2}$, ..., $(C_{lm})_{A_\kappa}$. These outgoing equivalent source strengths represent an efficient encoding of the pressure field for the scene. This step is repeated for all the elementary SH sources $s_{lm}(\mathbf{x},\mathbf{y})$ in the source representation (Equation 2), and the computed equivalent source strengths are stored. These equivalent sources are then used at runtime to compute pressure and its high-order derivatives.

Runtime We use the stored equivalent source strengths to compute the pressure fields $p_{lm}(\mathbf{x})$ for the elementary SH sources $s_{lm}(\mathbf{x})$:

$$p_{lm}(\mathbf{x}) = \sum_{j=1}^{\kappa} (C_{lm})^{tr}_{A_j} \Phi^{out}_{A_j}(\mathbf{x}) + s_{lm}(\mathbf{x}), \tag{9}$$

where $\Phi^{out}(\mathbf{x})$ is the equivalent source basis functions (as defined in Mehra et al. [16]). The total pressure field at the listener position due to the directional source $s(\mathbf{x},\mathbf{y})$ is computed as before. The derivative of the pressure field due to elementary source $\partial p_{lm}(\mathbf{x})/\partial x$ is computed by analytically differentiating the functions involved: the equivalent source basis functions $\Phi^{out}(\mathbf{x})$ and the source field $s_{lm}(\mathbf{x})$.

$$\frac{\partial p_{lm}(\mathbf{x})}{\partial x} = \sum_{j=1}^{\kappa} (C_{lm})^{tr}_{A_j} \frac{\partial \Phi^{out}_{A_j}(\mathbf{x})}{\partial x} + \frac{\partial s_{lm}(\mathbf{x})}{\partial x}. \tag{10}$$

The derivative of the total pressure field is computed as before.

6 Implementation

We use the 64-bit FASTBEM implementation of the boundary element method (`www.fastbem.com`). For ESM, the preprocessing code is implemented in MATLAB. The runtime code is implemented in C++ and has been integrated with Valve's Source game engine. The timing results for the BEM and ESM precomputation are measured on a 64-node CPU cluster (Xeon 5560 processor nodes, 8 cores, 2.80 GHz, 48 GB memory). The precomputation for each frequency is performed in parallel over all the nodes of the cluster. The timing results of our runtime system are measured on a single core of a 4-core 2.80 GHz Xeon X5560 desktop with 4 GB of RAM and NVIDIA GeForce GTX 480 GPU with 1.5 GB memory. Acoustic responses are computed up to the maximum simulation frequency of 1 kHz.

Measurement data The source directivity data used in our system is extracted from real-world measurement data provided by Meyer et al. [26]. The data is magnitude-only, averaged over the frequencies in each octave band, and is provided for all the octave bands within the frequency range of the sound sources. We use SH order $L = 3-4$ for the source representation, resulting in error less than $10-15\%$, which is a typical error threshold used for auralization purposes in interactive applications [10, 28, 16]. Error is defined as $\sqrt{\frac{\sum_i \sum_j \|D(\theta_i,\phi_j) - D_{SH}(\theta_i,\phi_j)\|^2}{\sum_i \sum_j \|D(\theta_i,\phi_j)\|^2}}$, where D is the measured directivity data and D_{SH} is our SH-based source directivity representation (Section 3.1). For spatial sound, we use the HRTF dataset provided by Algazi et al. [1], particularly for the KEMAR dummy. We compute the SH coefficients of this HRTF using standard SH projection methods [7].

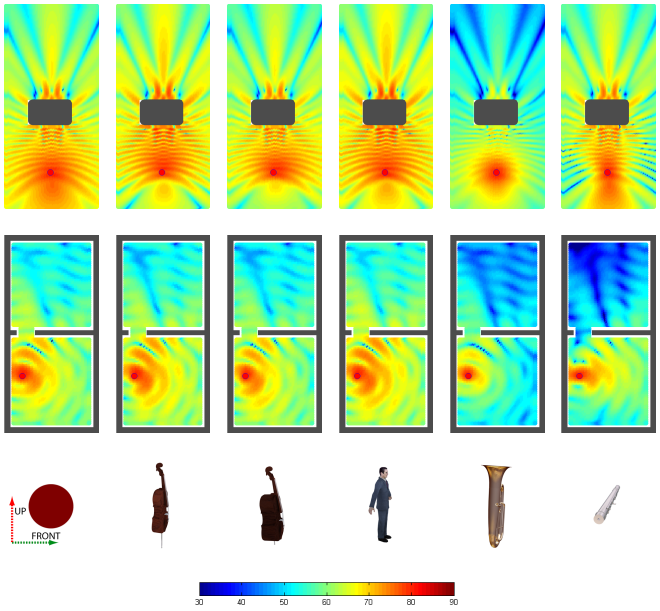


Fig. 3. **Directivity affects sound propagation**: Magnitude of pressure field (in dB) on a grid of listener positions for the single building and the empty room scene at 360 Hz with different directional sources. Source position is shown with a red dot. Front axis points towards the building in the single building and away from the left wall in the empty room. Up axis points upwards perpendicular to the listener grid.

HRTF modeling Our listener directivity approach supports any kind of SH-based HRTF models [27, 25]. In practice, we use the SH-based HRTF model proposed by Rafaely et al. [27]. A detailed evaluation to study the effect of HRTF SH order on spatial perception was conducted using various metrics such as inter-aural time difference (ITD), inter-aural level difference (ILD) and inter-aural cross-correlation coefficient (IACC) [27, 2]. These results indicate that a SH order of 1-2 is sufficient for spatial perception of frequencies up to 1 kHz; a SH order of 3 suffices up to 2 kHz; and a SH order of 3-6 suffices up to 8 kHz. Higher SH orders result in better spatial resolution, but computing the higher-order derivatives of pressure field for plane-wave decomposition also increases the computational cost. Therefore, the SH order may be determined based on the performance-accuracy trade-off.

Rotating sources and dynamic directivity After rotation by an angle ϕ about the z axis, the SH coefficients of the directivity function can be obtained by multiplying the vector of unrotated SH coefficients by the blockwise-sparse matrix given in [7, p. 23]. Matrices can also be derived for more general rotations about arbitrary axes [7, p. 21-26]. For handling dynamic directivity at runtime, we use the Intel MKL library to solve the linear system corresponding to the SH decomposition of the directivity function, at interactive rates.

Moving sources In order to enable moving sources, pressure fields are computed for directional sources positioned at regularly sampled

Table 1. **Precomputation cost:** Abbreviations are as follows- "#objs." number of ESM objects in the scene, "#srcs" number of directional sound sources, "#freq." number of frequency samples in the range (0-1 kHz), "L" is the SH order, and "ESM-sim" or "BEM-sim" is the total wall clock time to compute pressure fields using the ESM or BEM propagation techniques, respectively. The sound-propagation computations are performed in parallel for all the elementary SH sources for all the frequencies on a 64-node cluster with 8 cores per node. 'Crowd' and 'Music' scenes have no precomputation time, since only free-space propagation is performed.

ESM scenes	air volume	surface area	#objs. /#srcs	# freq. / L	ESM-sim (wall-clk)
Crowd	85^3m^3	–	0/1	250/4	–
Music	85^3m^3	–	0/1	250/3	–
Rotating	85^3m^3	$71m^2$	1/1	250/3	165 min
Parallel	85^3m^3	$142m^2$	2/1	250/3	195 min
Amphitheater	180^3m^3	$220m^2$	1/3	500/4	305 min
Reservoir	180^3m^3	$950m^2$	5/2	500/3	829 min
Christmas	180^3m^3	$2952m^2$	5/1	500/3	894 min

BEM scenes	air volume	surface area	#srcs	# freq. / L	BEM-sim (wall-clk)
Empty room	$24m^3$	$60m^2$	1	250/4	45 min
Furnished room	$66m^3$	$142m^2$	1	250/4	700 min

3D positions. In practice, a resolution of 1-2 meters with interpolation generates good results [28]. The sampling resolution is guided by the size of the human head and runtime memory requirements. This memory overhead increases linearly with the number of sampled source positions, which scales linearly with the size of the scene. Perceptual compression techniques, similar to those used in Raghuvanshi et al. [28], can be used to further reduce the overall memory cost.

Auralization Pressure fields for elementary SH sources, precomputed using the equivalent source technique and the SH coefficients of the HRTF, are loaded into Valve's game engine upon startup. At runtime, the acoustic responses of the elementary SH sources are evaluated and extrapolated to the output frequency (22 kHz), similar to [16]. SH coefficients of the source directivity are computed and used to generate the total pressure field and its derivatives at the listener position. As the listener (player) moves through the scene, the computed pressure and pressure derivatives are combined with the SH coefficients of the HRTF to compute binaural frequency responses, producing spatial sound at the listener. These frequency-response computations are performed asynchronously from both the visual-rendering pipeline and the audio-processing pipeline. As shown in Table 2, our technique can update binaural frequency responses at a rate of 10-15 Hz or more, which is sufficient for audio applications [9], while the game itself is able to perform visual rendering at 60 frames per second (or more). Audio processing is performed using FMOD (with Fourier transforms computed using Intel MKL), and rendered in frames of 1024 samples.

7 User Evaluation

We have conducted a preliminary user evaluation to study the effect of directivity on the realism of audio and audio-visual correlation in virtual environments. In this study, we compared the sound generated by a wave-based sound propagation technique without directivity, called the *base* method, and a wave-based sound propagation technique with our directivity, called *our* method. For this experiment, we chose the original ESM technique [16] as the *base* method and the ESM technique integrated with our directivity formulation (Section 5.2) as *our* method.

Study Design The study was designed for four comparison cases: *base vs. base*, *our vs. our*, *base vs. our* and *our vs. base*, which were tested over three benchmark scenes: parallel walls, reservoir, and rotating. For each comparison case and each benchmark scene, a pair of videos is generated with identical visuals but different audio as computed by the two sound propagation techniques in the comparison case. In total, twelve such video pairs are generated (4 comparison cases x 3 benchmarks). The study was conducted in form of an online survey where every subject was shown the 12 video pairs in random order. For each video pair, users were asked two questions: (a) "*Which of the two videos contains more realistic audio?*" and (b) "*Which of the two videos contains audio that matches better with the visuals shown?*". Our research hypotheses were: 1) Sound produced by *our* method will result in more realistic audio and better audio-visual correlation than that produced by the *base* method. 2) The level of increase in realism and audio-visual correlation for *our* method as compared to the *base* method depends upon the type of scene.

Procedure The study was conducted for a total of 43 subjects, all between the age of 18 and 48, made up of 27 males and 16 females. The mean age of the group is 27.8, and all of them had normal hearing and normal or corrected-to-normal vision. The average number of hours per week the subjects listened to music was 16.5. Before starting the study, the subjects were given detailed instructions and filled out a background survey questionnaire. Since the videos contained stereo audio, the subjects were required to use either headphones or earphones. The subjects were also asked to make sure they were wearing the headphones/earphones in correct orientation (left channel in left ear and so on). The next stage was the volume calibration session, in which the subjects were asked to play a test audio clip and change the volume of the audio system until the audio clip is barely audible. Subjects then started the study, in which they were presented with 12 video pairs and asked to rate each pair on a scale of 1 to 11 on two separate questions. A 1 response meant a strong preference for the first clip; 6 meant equal preference; and 11 meant a strong preference for the second clip. The subjects were allowed as much time as needed and were free to take a break at any time. The study concluded after the subject finished rating all the 12 videos. After completing the study, the subjects were encouraged to give feedback and thanked for their time and efforts.

8 Results

In Table 1, we show the cost of precomputing the pressure fields using BEM and ESM techniques. The three independent computations (for all sound sources; the elementary SH sources; and the different frequencies) can be easily performed in parallel. Table 2 shows the runtime performance and memory requirements of the ESM technique for computing the pressure field due to directional source at a moving listener. BEM computes the pressure field for the empty room and the furnished room at the listener position in 3 and 5 seconds, respectively. The storage cost for the BEM pressure fields is 716 MB and 1.4 GB for the empty and furnished room, respectively. BEM has a much higher memory requirement than the ESM (Table 4 in [16]) since it captures the sound-field interactions close to the surface of the object. ESM, on the other hand, captures the sound-field interactions outside the object's offset surface. Note that sound propagation is inherently a much more memory-intensive task than visual rendering due to sound's broad frequency range, the complex phases of the sound fields, and its non-diffuse reflections and wave effects.

Figure 3 demonstrates how directivity affects sound propagation. Figure 4 shows that with increasing frequency of sound sources, we need increasingly higher-order terms in the SH representation. As a general trend, sound source directivity becomes sharper (more prominent) with increasing frequency, requiring higher-order SH basis functions. Refer to the supplementary video for auralizations of different scenes.

8.1 Validation

We present validation results for our SH directivity and sound propagation techniques.

Source modeling For SH decomposition of the directivity function, we did not observe any significant ringing for the first four oc-

Table 2. **ESM runtime performance:** "L" is the SH order, and "# eq. srcs" is the number of equivalent sources (in millions). For each scene, the runtime performance "eval." with and without spatial sound and the storage requirement "storage" are shown. Storage numbers indicate fixed cost for equivalent source positions and the total cost for storing the strengths for all elementary SH sources up to SH order L. 'Crowd' and 'Music' scenes have no equivalent sources, since only free-space propagation is performed.

Scene	L	# eq. srcs	eval. (no spatial, per src)	eval. (spatial, per src)	storage (total, fixed + per src)
Crowd	4	–	0.14 ms	0.34 ms	–
Music	3	–	0.14 ms	0.34 ms	–
Rotating	3	0.1 M	12.6 ms	16.7 ms	(1 + 29) MB
Parallel	3	0.2 M	17.9 ms	24 ms	(2 + 58) MB
Amphitheater	4	0.1 M	14 ms	18.6 ms	(1 + 51) MB
Reservoir	3	0.8 M	86 ms	115 ms	(10 + 230) MB
Christmas	3	0.7 M	90.6 ms	119.5 ms	(8 + 202) MB

tave bands. For the last two octave bands, ringing was resolved by the standard technique of windowing the truncated SH coefficients using Lanczos sigma factors [29]. Our SH-based source representation converges to the measured directivity with increasing SH order (see Figure 5).

Sound propagation To validate our technique's modeling of source directivity's effect on sound propagation, we performed validation experiments against the Biot-Tolstoy-Medwin (BTM) method [32, 15]. This is an offline technique that provides a wideband reference solution with accurate diffraction in simple scenes and can incorporate source directivities. We integrated our data-driven SH-based directivity formulation with the BTM Toolbox for MATLAB (`www.iet.ntnu.no/~svensson/software/index.html#EDGE`). We performed BTM simulations over three fixed-receiver positions and a fixed source position in the wall scene. Simulations for receivers 1, 2, and 3 took 57, 27, and 18 minutes, respectively. Please refer to the supplementary video for receiver positions and corresponding auralizations.

Integration with frequency-domain techniques The pressure field in frequency domain is a complex quantity with magnitude and phase. Figure 6 shows the magnitude of the pressure field computed by the BEM simulation and the ESM technique for directional sources. BEM error tolerance and order of multipole expansion was set to 1% and 6, respectively. ESM error thresholds for scattering and interactive matrices are 15% and 1%, respectively. ESM error thresholds were chosen for interactive performance and can be reduced to achieve higher accuracy. The pressure fields produced by the two techniques agree within error of $< 5-10\%$. The offline BEM method has a much higher computational and memory overhead, but the resulting pressure fields are more accurate.

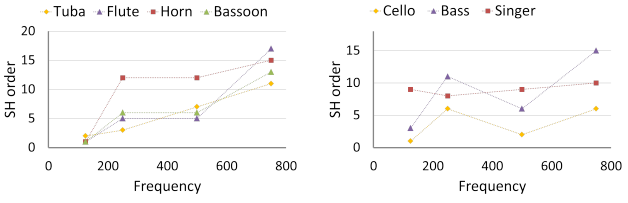
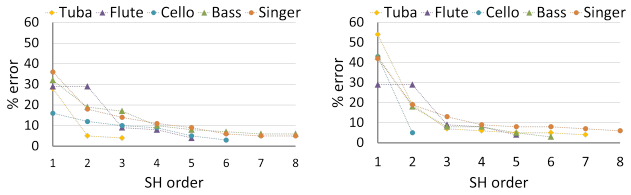
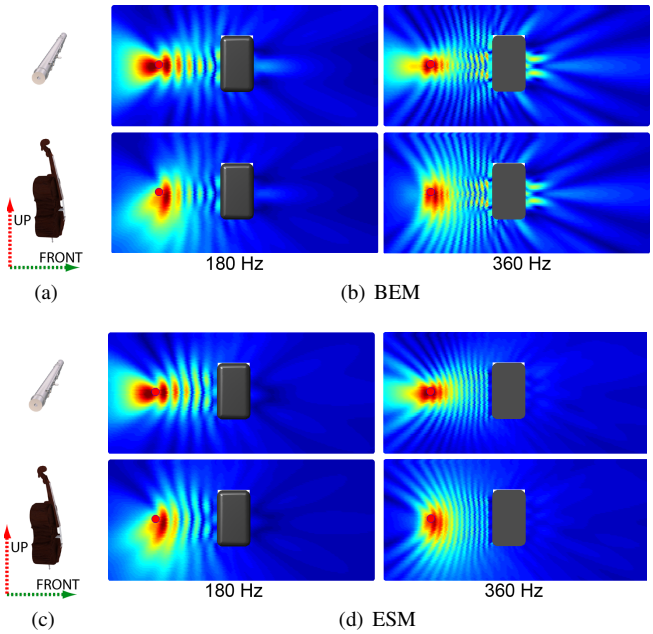


Fig. 4. **Growth**: Directivity of sound sources becomes sharper with frequency. To approximate the directivity function at higher frequencies while maintaining a constant error threshold of 5%, higher-order SH source representation is required.



(a) 177-354 Hz (b) 355-710 Hz

Fig. 5. **Convergence**: Our SH-based source representation converges to the measured directivity data with higher SH orders, for different directional sources. Error metric defined in Section 6.



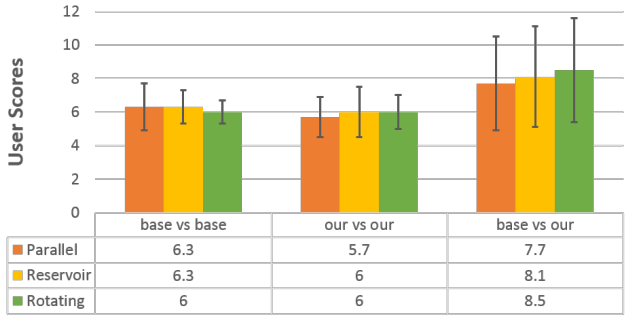
(a) (b) BEM (c) (d) ESM

Fig. 6. **Integration of source directivity with frequency-domain techniques:** Magnitude of pressure field (in Pa, normalized $[0,1]$) using offline BEM and interactive ESM technique for the single wall scene at 180 Hz and 360 Hz. Source position is shown with a red dot. Front axis points towards the building. Up axis points upwards perpendicular to the listener grid.

8.2 Evaluation

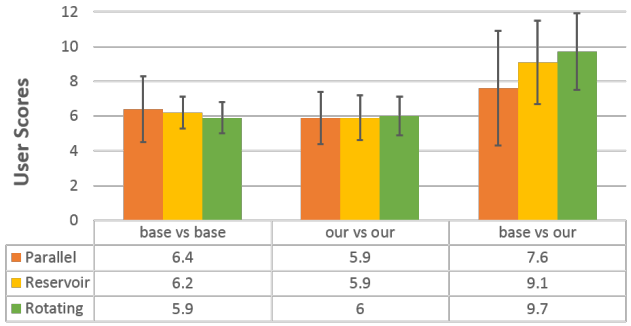
Figure 7 shows the mean and standard deviation of the subjects' scores for the two questions. The scores of *our vs. base* comparison case were reversed and combined it with *base vs. our*. For the *base vs. base* and *our vs. our* case, the subjects showed equal preference, as expected, exemplified by the mean scores ranging from 5.7 to 6.4. However, for the *base vs. our* case, the subjects showed an increasing preference for *our* method as exemplified by the higher mean scores (ranging from 7.6 to 9.7). This is consistent with hypothesis 1. In addition, the mean scores of *base vs. our* case increased from parallel to the rotating scene-type, consistent with hypothesis 2. We hypothesize that in the parallel wall scene, which contains multiple propagation effects (diffraction low-passing, moving source, rotating listener, source and listener directivity) all happening at the same time, the effect of directivity was masked out to some extent. In the rotating scene, source and listener directivity were the dominant propagation effects present in the scene, resulting in comparatively higher scores.

The scores were analysed with a two-way analysis of variance (ANOVA). For the question regarding audio-visual correlation, we found significant effects for both the comparison case ($F(2,84)=67.36$, $p<0.0001$) and scene-type ($F(2,84)=3.72$, $p=0.0284$). A significant in-



	base vs base	our vs our	base vs our
Parallel	6.3	5.7	7.7
Reservoir	6.3	6	8.1
Rotating	6	6	8.5

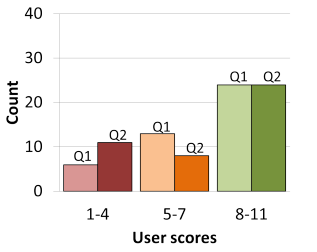
(a) "Level of realistic audio"



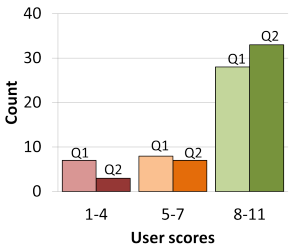
	base vs base	our vs our	base vs our
Parallel	6.4	5.9	7.6
Reservoir	6.2	5.9	9.1
Rotating	5.9	6	9.7

(b) "Audio-visual correlation"

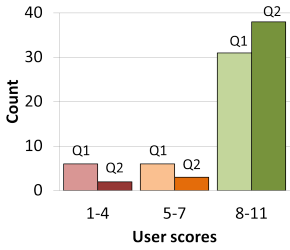
Fig. 7. **User study:** Mean and standard deviation of subject's responses are tabulated for the two questions regarding "realistic audio" and "audio-visual correlation". The term **base** refers to the wave-based sound propagation without directivity; **our** refers to wave-based propagation with our directivity approach integrated.



(a) Parallel



(b) Reservoir



(c) Rotating

Fig. 8. **Histogram of subject scores for the comparison between the *base* method (without directivity) and *our* method (with directivity):** Q1 and Q2 refers to the question regarding the level of realistic audio and the level of audio-visual correlation respectively.

teraction between the case and scene-type factors was found as well ($F(4,168)=8.18$, $p<0.0001$). This can be explained by Figure 7(b) where the variation of the scores with scene-type changes for different comparison-cases. For the question regarding realistic audio, we found significant effect for the comparison case ($F(2,84)=35.78$, $p<0.0001$). However, we did not find any significant effect for the scene-type ($F(2,84)=0.86$, $p=0.42$) and the interaction between case and scene-type ($F(4,168)=1.05$, $p=0.38$). In this case, even though the scores for the *base vs. our* case is higher than other two cases, the variation of the scores with scene-type does not change as significantly as before for different comparison cases.

In Figure 8, we plot the histograms of the scores for the *base vs. our* case for the three scenes. The scores are distributed in three bins: (1-4) means preference for *base* method, (5-7) means equal preference and (8-11) means preference for *our* method. For the question regarding realism, 56%, 65% and 76% of the participants preferred *our* method for the three scenes respectively. Furthermore, for audio-visual correlation, 56%, 76% and 88% of the participants preferred *our* method. These results show that majority of subjects found the sound generated using our directivity approach to be more realistic and correlated better with the visuals as compared to sound generated without directivity. This is a preliminary user study and we plan to conduct a more extensive user evaluation of our technique in the future.

9 Conclusion and Future Work

We have presented a novel approach for incorporating dynamic source and listener directivity in a general frequency-domain, wave-based sound propagation technique. Our approach can automatically model wave-effects from low-frequency directional sources and can handle analytical, data-driven, rotating or dynamic directivity at runtime. We have also described an efficient plane-wave decomposition approach to handle HRTF-based listener directivity in order to perform spatial sound rendering at interactive rates.

Directional sources with sharp directivity patterns (delta functions) would be expensive to handle with our current approach since it would require a very high-order SH expansion. We want to explore other basis functions, such as wavelets, to handle sharp directivities. We would also like to add support for artist-controlled directivity, allowing real-time feedback on the effect of directivity on propagated sound. We have used magnitude-only directivity data in our current implementation, but our approach can easily support complex data (magnitude and phase), which we plan to test in the future. Our source formulation can handle both near- and far-field sound radiation by directional sources (equation 1). However, near-field directivity requires a set of dense measurements of complex frequency responses very close to the source at twice the Nyquist rate, which is currently unavailable. We hope that such a dataset becomes available in the future. Lastly, since wave-based approaches are computationally limited to a few kHz frequency, we would like to explore hybridization of wave-based techniques with geometric approaches to handle directional sources over the complete audible frequency range.

Acknowledgments

The authors wish to thank Anish Chandak, John Snyder, Nikunj Raghuvanshi and Zhimin Ren for simulating discussions, and the anonymous reviewers for their valuable feedback. We would like to acknowledge Ming Lin and anonymous users for help with the user study and evaluation. We would also like to thank Valve Corporation for permission to use the Source SDK and Half-Life 2 artwork for the Reservoir and Parallel walls scenes. This work was supported by Link Foundation Fellowship in Advanced Simulation and Training, ARO Contract W911NF-10-1-0506, W911NF-12-1-0430, W911NF-13-C-0037, and NSF awards 0917040 and 1320644.

References

[1] V. Algazi, R. Duda, D. Thompson, and C. Avendano. The CIPIC HRTF database. In *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the*, pages 99 –102, 2001.

[2] A. Avni. Spaciousness of sound fields captured by spherical microphone arrays. Master's thesis, Bengurion University of the Negev Faculty of Engineering Sciences, 2010.

[3] D. R. Begault. *3D Sound for Virtual Reality and Multimedia*. Academic Press, 1994.

[4] J. Blauert. *Spatial hearing: the psychophysics of human sound localization*. Mit Press, 1983.

[5] J. N. Chadwick, S. S. An, and D. L. James. Harmonic shells: a practical nonlinear sound model for near-rigid thin shells. NY, USA, 2009. ACM.

[6] T. Funkhouser, N. Tsingos, and J.-M. Jot. Survey of methods for modeling sound propagation in interactive virtual environment systems. *Presence and Teleoperation*, 2003.

[7] R. Green. Spherical Harmonic Lighting: The Gritty Details. *Archives of the Game Developers Conference*, Mar. 2003.

[8] H. Hacihabiboglu, B. Gunel, and A. Kondoz. Time-domain simulation of directive sources in 3-d digital waveguide mesh-based acoustical models. *Audio, Speech, and Language Processing, IEEE Transactions on*, 16(5):934–946, 2008.

[9] IASIG. Interactive audio special interest group : Interactive 3d audio rendering guidelines, level 2.0, 1999.

[10] D. L. James, J. Barbič, and D. K. Pai. Precomputed acoustic transfer: output-sensitive, accurate sound generation for geometrically complex vibration sources. In *ACM SIGGRAPH 2006 Papers*, SIGGRAPH '06, pages 987–995, New York, NY, USA, 2006. ACM.

[11] H. Jers. Directivity of singers. *The Journal of the Acoustical Society of America*, 118(3):2008–2008, 2005.

[12] G. Kino. *Acoustic waves: devices, imaging, and analog signal processing*. Prentice-Hall Signal Processing Series. Prentice Hall PTR, 1987.

[13] P. Larsson, D. Vastfjall, and M. Kleiner. Better presence and performance in virtual environments by improved binaural sound rendering. In *Virtual, Synthetic, and Entertainment Audio conference*, Jun 2002.

[14] Y. Liu. *Fast Multipole Boundary Element Method: Theory and Applications in Engineering*. Cambridge University Press, 2009.

[15] T. Lokki, P. Svensson, and L. Savioja. An efficient auralization of edge diffraction. In *Architectural Acoustics and Sound Reinforcement Conf.*, Jun 2002.

[16] R. Mehra, N. Raghuvanshi, L. Antani, A. Chandak, S. Curtis, and D. Manocha. Wave-based sound propagation in large open scenes using an equivalent source formulation. *ACM Trans. Graph.*, Apr. 2013.

[17] M. Menzies, Dylan; Al-akaidi. Ambisonic synthesis of complex sources. *J. Audio Eng. Soc*, 55(10):864–876, 2007.

[18] J. Meyer and U. Hansen. *Acoustics and the Performance of Music (Fifth edition)*. Lecture Notes in Mathematics. Springer, 2009.

[19] N. Noisternig and B. Katz. Reconstructing sound source directivity in virtual acoustic environments. In *International workshop on the Principles and Applicaions of Spatial Hearing (IWPASH)*, 2009.

[20] M. Ochmann. The source simulation technique for acoustic radiation problems. *Acustica*, 81:512–527, 1995.

[21] M. Ochmann and F. Mechel. *Analytical and Numerical Methods in Acoustics*. Springer Berlin Heidelberg, 2004.

[22] F. Otondo and J. H. Rindel. The influence of the directivity of musical instruments in a room. *Acta Acustica*, 90(6):1178–1184, 2004.

[23] M. Park and B. Rafaely. Sound-field analysis by plane-wave decomposition using spherical microphone array. *The Journal of the Acoustical Society of America*, 118(5):3094–3103, 2005.

[24] A. D. Pierce. *Acoustics: An Introduction to Its Physical Principles and Applications*. Acoustical Society of America, 1989.

[25] M. Pollow, K.-V. Nguyen, O. Warusfel, T. Carpentier, M. Muller-Trapet, M. Vorlander, and M. Noisternig. Calculation of head-related transfer functions for arbitrary field points using spherical harmonics decomposition. *Acta Acustica united with Acustica*, 98(1):72–82, 2012.

[26] PTB. http://www.ptb.de/cms/en/fachabteilungen/abt1/fb-16/ag-1630/room-acoustics/directivities.html (last viewed Oct. 23, 2012). 1978.

[27] B. Rafaely and A. Avni. Interaural cross correlation in a sound field represented by spherical harmonics. *The Journal of the Acoustical Society of America*, 127(2):823–828, 2010.

[28] N. Raghuvanshi, J. Snyder, R. Mehra, M. C. Lin, and N. K. Govindaraju. Precomputed Wave Simulation for Real-Time Sound Propagation of Dynamic Sources in Complex Scenes. *SIGGRAPH 2010*, 29(3), July 2010.

[29] P.-P. Sloan. Stupid spherical harmonics tricks. In *Game Developers Conference*, 2008.

[30] A. Southern and D. Murphy. Low complexity directional sound sources for finite difference time domain room acoustic models. In *Audio Engineering Society Convention 126*, 5 2009.

[31] R. L. Storms and M. Usa. Auditory-visual cross-modal perception. In *In ICAD*, 2000.

[32] U. P. Svensson, R. I. Fred, and J. Vanderkooy. An analytic secondary source model of edge diffraction impulse responses . *Acoustical Society of America Journal*, 106:2331–2344, Nov. 1999.

[33] M. C. Vigeant. Investigations of incorporating source directivity into room acoustics computer models to improve auralizations. *The Journal of the Acoustical Society of America*, 124(5):2664–2664, 2008.

[34] M. Vorlander. Simulation of the transient and steady-state sound propagation in rooms using a new combined ray-tracing/image-source algorithm. *The Journal of the Acoustical Society of America*, 86(1):172–178, 1989.

[35] O. Warusfel and N. Misdariis. Sound source radiation syntheses: From performance to domestic rendering. In *Audio Engineering Society Convention 116*, May 2004.

[36] D. N. Zotkin, R. Duraiswami, and N. A. Gumerov. Plane-wave decomposition of acoustical scenes via spherical and cylindrical microphone arrays. *IEEE Trans. Audio, Speech and Language Processing*, 18, 2010.