Rigid sphere room impulse response simulation: Algorithm and applications

D. P. Jarrett[a)]
Communications and Signal Processing Group, Department of Electrical & Electronic Engineering, Imperial College London, Exhibition Road, London SW7 2AZ, United Kingdom

E. A. P. Habets
International Audio Laboratories Erlangen, Am Wolfsmantel 33, D-91058 Erlangen, Germany

M. R. P. Thomas[b)] and P. A. Naylor
Communications and Signal Processing Group, Department of Electrical & Electronic Engineering, Imperial College London, Exhibition Road, London SW7 2AZ, United Kingdom

Simulated room impulse responses have been proven to be both useful and indispensable for comprehensive testing of acoustic signal processing algorithms while controlling parameters such as the reverberation time, room dimensions, and source–array distance. In this work, a method is proposed for simulating the room impulse responses between a sound source and the microphones positioned on a spherical array. The method takes into account specular reflections of the source by employing the well-known image method, and scattering from the rigid sphere by employing spherical harmonic decomposition. Pseudocode for the proposed method is provided, taking into account various optimizations to reduce the computational complexity. The magnitude and phase errors that result from the finite order spherical harmonic decomposition are analyzed and general guidelines for the order selection are provided. Three examples are presented: an analysis of a diffuse reverberant sound field, a study of binaural cues in the presence of reverberation, and an illustration of the algorithm's use as a mouth simulator. © *2012 Acoustical Society of America.*
[http://dx.doi.org/10.1121/1.4740497]

I. INTRODUCTION

In general, the evaluation of acoustic signal processing algorithms, such as source localization and speech enhancement algorithms, makes use of simulated room transfer functions (RTFs). By using simulated RTFs it is possible to comprehensively evaluate an algorithm under many acoustic conditions (e.g., reverberation time, room dimensions, and source–array distance). Allen and Berkley's image method[1] is a widely used approach to simulate RTFs between an omnidirectional sound source and one or more microphones in a reverberant environment. In the last few decades, several extensions have been proposed.[2,3]

In recent years the use of spherical microphone arrays has become prevalent, due to their ability to analyze sound fields in three dimensions.[4,5] These arrays are commonly one of two types: the open array, where microphones are suspended in free space on an "open" sphere, and the rigid array, where microphones are mounted on a rigid baffle. The rigid sphere is often preferred as it improves the numerical stability of many processing algorithms[6] and its scattering effects are rigorously calculable.[5]

Currently, many works relating to spherical array processing consider only free-field responses, however, when a rigid array is used, the rigid baffle causes scattering of the sound waves incident upon the array that the image method does not consider. This scattering has an effect on the room transfer functions, especially at high frequencies and/or for microphones situated on the occluded side of the array. Furthermore, the reverberation due to room boundaries, such as walls, ceiling, and floor, must also be considered, particularly in small rooms.

Although measured transfer functions include both these effects, they are both time-consuming and expensive to acquire. A method for simulating RTFs in a reverberant room, while accounting for scattering, is therefore essential, allowing for fast, comprehensive, and repeatable testing. We propose such a method that combines a model of the scattering in the spherical harmonic domain with a version of the image method that accounts for reverberation in a computationally efficient way.

The simulated RTFs include the direct path, reflections due to room reverberation, scattering of the direct path and scattering of the reverberant reflections. Reflections of the scattered sound and multiple interactions between the room boundaries and the sphere are excluded as they do not contribute significantly to the sound field, provided the distances between the room boundaries and the sphere are several times the sphere's radius,[7] which is easily achieved in the case of a small scatterer.[8]

---

a)Author to whom correspondence should be addressed. Electronic mail: daniel.jarrett05@imperial.ac.uk
b)Present address: Microsoft Research, One Microsoft Way, Redmond, WA 98052.

Furthermore, we assume an empty rectangular shoebox room (with the exception of the rigid sphere) and specular reflections, as was assumed in the conventional image method.[1] Finally, the scattering model used assumes a perfectly rigid baffle, without absorption.

In this paper, we first briefly summarize Allen and Berkley's image method and then present our proposed method in the spherical harmonic domain. We then discuss some implementation aspects, namely, the truncation of an infinite sum in the room transfer function expression and the reduction of the method's computational complexity, and then provide a pseudocode description of the method. An open-source software implementation is available online.[9] Finally, we show some example uses of the method and, where possible, compare the simulated results obtained with theoretical models.

An early version of part of this work was initially presented in Ref. 10. The novel contributions of this current paper include an additional analysis of the scattering model used to model the rigid sphere, reference to experimental work that validates this model, a more comprehensive presentation of the proposed method and how it differs from the conventional image method, a discussion of computational complexity, pseudocode for the proposed method, and a number of examples and applications based on the proposed method.

II. ALLEN AND BERKLEY'S IMAGE METHOD

The source-image or image method[1] is one of the most commonly used room acoustics simulation methods in the acoustic signal processing community. The principle of the method is to model a RTF as the sum of a direct path component and a number of discrete reflections, each of these components being represented in the RTF by a free-space Green's function. In this section, we review the free-space Green's function and the image method.

A. Green's function

For a source at a position $\mathbf{r}_s$ and a receiver at a position $\mathbf{r}$, the free-space Green's function, a solution to the inhomogeneous Helmholtz equation applying the Sommerfeld radiation condition, is given by[11]

$$G(\mathbf{r}|\mathbf{r}_s, k) = \frac{e^{+ik\|\mathbf{r}-\mathbf{r}_s\|}}{4\pi\|\mathbf{r}-\mathbf{r}_s\|}, \tag{1}$$

where $\|\cdot\|$ denotes the ℓ-2 norm and the wavenumber k is related to frequency f (Hz) and the speed of sound c (m/s) via the relationship $k = 2\pi f/c$.

In the time-domain, the Green's function is given by

$$g(\mathbf{r}|\mathbf{r}_s, t) = \frac{\delta\left(t - \frac{\|\mathbf{r}-\mathbf{r}_s\|}{c}\right)}{4\pi\|\mathbf{r}-\mathbf{r}_s\|}, \tag{2}$$

where δ is the Dirac delta function and t is time. This corresponds to a pure impulse at time $t = \frac{\|\mathbf{r}-\mathbf{r}_s\|}{c}$, i.e., the propagation time from $\mathbf{r}_s$ to $\mathbf{r}$.

B. Image method

Consider a rectangular room with length L_x, width L_y, and height L_z. The reflection coefficients of the four walls, floor, and ceiling are β_{x_1}, β_{x_2}, β_{y_1}, β_{y_2}, β_{z_1} and β_{z_2}, where the v_1 coefficients ($v \in \{x, y, z\}$) correspond to the boundaries at $v = 0$ and the v_2 coefficients correspond to the boundaries at $v = L_v$.

If the sound source is located at $\mathbf{r}_s = (x_s, y_s, z_s)$ and the receiver is located at $\mathbf{r} = (x, y, z)$, the images obtained using the walls at $x = 0$, $y = 0$, and $z = 0$ can be expressed as a vector $\mathbf{R_p}$,

$$\mathbf{R_p} = [x_s - x + 2p_x x, y_s - y + 2p_y y, z_s - z + 2p_z z], \tag{3}$$

where each of the elements in $\mathbf{p} = (p_x, p_y, p_z)$ can take values 0 or 1, thus resulting in eight combinations that form a set $\mathcal{P}$. To consider all reflections we also define a vector $\mathbf{R_m}$, which we add to $\mathbf{R_p}$,

$$\mathbf{R_m} = [2m_x L_x, 2m_y L_y, 2m_z L_z], \tag{4}$$

where each of the elements in $\mathbf{m} = (m_x, m_y, m_z)$ can take values between $-N_m$ and N_m, and N_m is used to limit computational complexity and circular convolution errors, thus resulting in a set $\mathcal{M}$ of $(2N_m + 1)^3$ combinations. The image positions in the x and y dimensions are illustrated in Fig. 1.
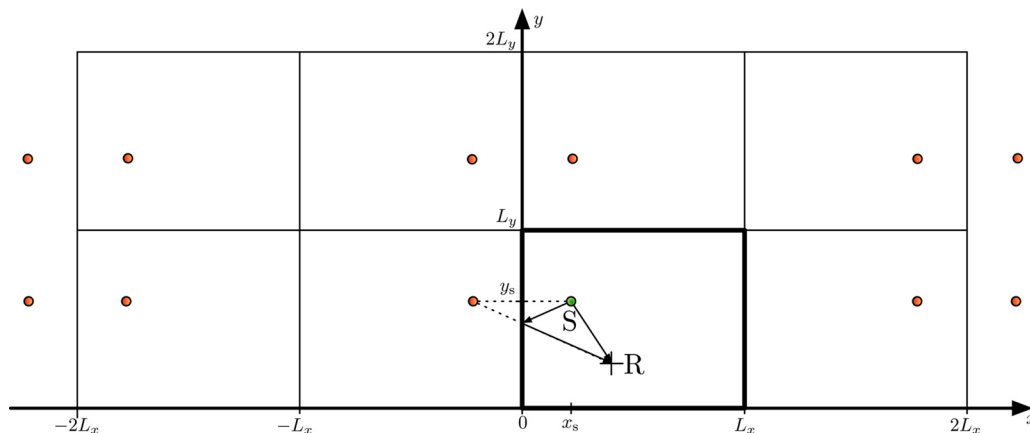

FIG. 1. (Color online) A slice through the image space showing the positions of the images in the x and y dimensions, with a source S and receiver R. The full image space has three dimensions (x, y, and z). An example of a reflected path (first order reflection about the x-axis) is also shown.

Jarrett *et al.*: Rigid sphere room impulse response simulation   1463

The distance between an image and the receiver is given by $\|\mathbf{R_p} + \mathbf{R_m}\|$. Using Eq. (1), the RTF H is then given by

$$H(\mathbf{r}|\mathbf{r_s}, k) = \sum_{\mathbf{p} \in \mathcal{P}} \sum_{\mathbf{m} \in \mathcal{M}} \beta_{x_1}^{|m_x+p_x|} \beta_{x_2}^{|m_x|} \beta_{y_1}^{|m_y+p_y|} \beta_{y_2}^{|m_y|} \beta_{z_1}^{|m_z+p_z|} \beta_{z_2}^{|m_z|}$$
$$\times \frac{e^{+ik(\|\mathbf{R_p}+\mathbf{R_m}\|)}}{4\pi\|\mathbf{R_p}+\mathbf{R_m}\|}. \quad (5)$$

Using Eq. (2), we obtain the room impulse response (RIR)

$$h(\mathbf{r}|\mathbf{r_s}, t) = \sum_{\mathbf{p} \in \mathcal{P}} \sum_{\mathbf{m} \in \mathcal{M}} \beta_{x_1}^{|m_x+p_x|} \beta_{x_2}^{|m_x|} \beta_{y_1}^{|m_y+p_y|} \beta_{y_2}^{|m_y|} \beta_{z_1}^{|m_z+p_z|} \beta_{z_2}^{|m_z|}$$
$$\times \frac{\delta\left(t - \frac{(\|\mathbf{R_p}+\mathbf{R_m}\|)}{c}\right)}{4\pi\|\mathbf{R_p}+\mathbf{R_m}\|}. \quad (6)$$

III. PROPOSED METHOD IN THE SPHERICAL HARMONIC DOMAIN

There exists a compact analytical expression for the scattering due to the rigid sphere in the spherical harmonic domain, therefore we first express the free-space Green's function in this domain, and then use this to form an expression for the RTF including scattering.

The proposed rigid sphere scattering model[12] has a long history in the literature; it was first developed by Clebsch and Rayleigh in 1871–1872.[13] It is presented in a number of acoustics texts,[14–16] and is used in many theoretical analyses for spherical microphone arrays.[17,18] In addition to being widely used in theory, this model has also been experimentally validated by Duda and Martens[19] using a microphone inserted in a hole drilled through a 10.9 cm radius bowling ball placed in an anechoic chamber. This is a reasonable approximation to a spherical microphone array; indeed a bowling ball was also used by Li and Duraiswami to construct a hemispherical microphone array.[20]

A. Green's function

We define position vectors in spherical coordinates relative to the center of our array. Letting r be the array radius and Ω an elevation-azimuth pair, the microphone position vector is defined as $\tilde{\mathbf{r}} \triangleq (r, \Omega)$ (where $\tilde{\ }$ indicates a vector in spherical coordinates). Similarly, the source position vector is given by $\tilde{\mathbf{r}}_s \triangleq (r_s, \Omega_s)$. It is hereafter assumed that where the addition, ℓ-2 norm or scalar product operations are applied to spherical polar vectors, they have previously been converted to Cartesian coordinates.

The free-space Green's function (1) can be expressed in the spherical harmonic domain using the following spherical harmonic decomposition (SHD):[16]

$$G(\tilde{\mathbf{r}}|\tilde{\mathbf{r}}_s, k) = \frac{e^{+ik\|\tilde{\mathbf{r}} - \tilde{\mathbf{r}}_s\|}}{4\pi\|\tilde{\mathbf{r}} - \tilde{\mathbf{r}}_s\|}$$
$$= ik \sum_{l=0}^{\infty} \sum_{m=-l}^{l} j_l(kr) h_l^{(1)}(kr_s) Y_{lm}^*(\Omega_s) Y_{lm}(\Omega)$$
$$= ik \sum_{l=0}^{\infty} j_l(kr) h_l^{(1)}(kr_s) \sum_{m=-l}^{l} Y_{lm}^*(\Omega_s) Y_{lm}(\Omega) \quad (7)$$

where Y_{lm} is the spherical harmonic function of order l and degree m, j_l is the spherical Bessel function of order l and $h_l^{(1)}$ is the spherical Hankel function of the first kind and of order l. This decomposition is also known as a spherical Fourier series expansion or spherical harmonics expansion of the Green's function.

According to the spherical harmonic addition theorem,[16]

$$\sum_{m=-l}^{l} Y_{lm}^*(\Omega_s) Y_{lm}(\Omega) = \frac{2l+1}{4\pi} P_l(\cos \Theta_{\tilde{\mathbf{r}}, \tilde{\mathbf{r}}_s}), \quad (8)$$

where P_l is the Legendre polynomial of degree l and $\Theta_{\tilde{\mathbf{r}}, \tilde{\mathbf{r}}_s}$ is the angle between $\tilde{\mathbf{r}}$ and $\tilde{\mathbf{r}}_s$. Using this theorem, which in many cases reduces the complexity of the implementation, we can simplify the Green's function in Eq. (7) to

$$G(\tilde{\mathbf{r}}|\tilde{\mathbf{r}}_s, k) = \frac{ik}{4\pi} \sum_{l=0}^{\infty} j_l(kr) h_l^{(1)}(kr_s)(2l+1) P_l(\cos \Theta_{\tilde{\mathbf{r}}, \tilde{\mathbf{r}}_s}). \quad (9)$$

The angle $\Theta_{\tilde{\mathbf{r}}, \tilde{\mathbf{r}}_s}$ is obtained as the dot product of the two normalized vectors $\hat{\mathbf{r}}_s = \tilde{\mathbf{r}}_s/r_s$ and $\hat{\mathbf{r}} = \tilde{\mathbf{r}}/r$:

$$\cos \Theta_{\tilde{\mathbf{r}}, \tilde{\mathbf{r}}_s} = \hat{\mathbf{r}} \cdot \hat{\mathbf{r}}_s. \quad (10)$$

B. Neumann Green's function

The free-space Green's function describes the propagation of sound in free space only. However, when a rigid sphere is present, a boundary condition must hold: the radial velocity must vanish on the surface of the sphere. The function $G_N(\tilde{\mathbf{r}}|\tilde{\mathbf{r}}_s, k)$ satisfying this boundary condition is called the *Neumann Green's* function, and describes the sound propagation between a point $\tilde{r}_s$ and a point $\tilde{r}$ on the rigid sphere,[16]

$$G_N(\tilde{\mathbf{r}}|\tilde{\mathbf{r}}_s, k) = G(\tilde{\mathbf{r}}|\tilde{\mathbf{r}}_s, k) - \frac{ik}{4\pi} \sum_{l=0}^{\infty} \frac{j_l'(kr)}{h_l^{(1)'}(kr)} h_l^{(1)}(kr) \quad \text{★}$$
$$\times h_l^{(1)}(kr_s)(2l+1) P_l(\cos \Theta_{\tilde{\mathbf{r}}, \tilde{\mathbf{r}}_s})$$
$$= \frac{ik}{4\pi} \sum_{l=0}^{\infty} b_l(kr) h_l^{(1)}(kr_s)(2l+1) P_l(\cos \Theta_{\tilde{\mathbf{r}}, \tilde{\mathbf{r}}_s}) \quad (11)$$

where $(\cdot)'$ denotes the first derivative and the term

$$b_l(kr) = j_l(kr) - \frac{j_l'(kr)}{h_l^{(1)'}(kr)} h_l^{(1)}(kr) \quad (12)$$

is often called the *mode strength*. For the open sphere, $b_l(kr) = j_l(kr)$ yields the free-space Green's function. The Wronskian relation[16]

$$j_l(x) h_l^{(1)'}(x) - j_l'(x) h_l^{(1)}(x) = \frac{i}{x^2} \quad (13)$$

allows us to simplify Eq. (12) to

$$b_l(kr) = \frac{i}{h_l^{(1)'}(kr)(kr)^2}. \quad (14)$$

Jarrett *et al.*: Rigid sphere room impulse response simulation

* Please note that line 1 of equation (11) has been modified above to correct a printing/typesetting error in the version of the paper published by JASA.
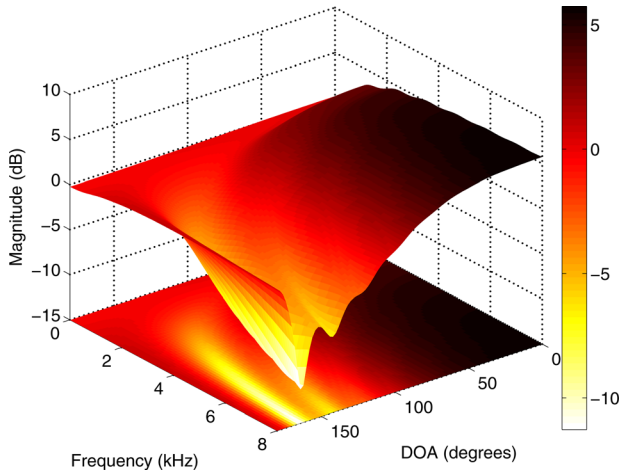
FIG. 2. (Color online) Magnitude of the response between a source and a receiver placed on a rigid sphere of radius 5 cm at a distance of 1 m, as a function of frequency and direction of arrival (DOA). The response was normalized with respect to the free-field response.

C. Scattering model behavior

The behavior of the scattering model is illustrated in Fig. 2, which plots the magnitude of the response between a source and a receiver on a rigid sphere of radius 5 cm for a source–array distance of 1 m, as a function of frequency and direction of arrival (DOA). The response was normalized using the free-field/open sphere response, therefore, the plot shows only the effect due to scattering. Due to rotational symmetry, we only looked at the one-dimensional DOA, instead of looking at both azimuth and elevation, and limited the DOA to the 0°–180° range.

When the source is located on the same side of the sphere as the receiver (i.e., the direction of arrival is 0°), the rigid sphere response is greater than the open sphere response due to constructive scattering, tending toward a
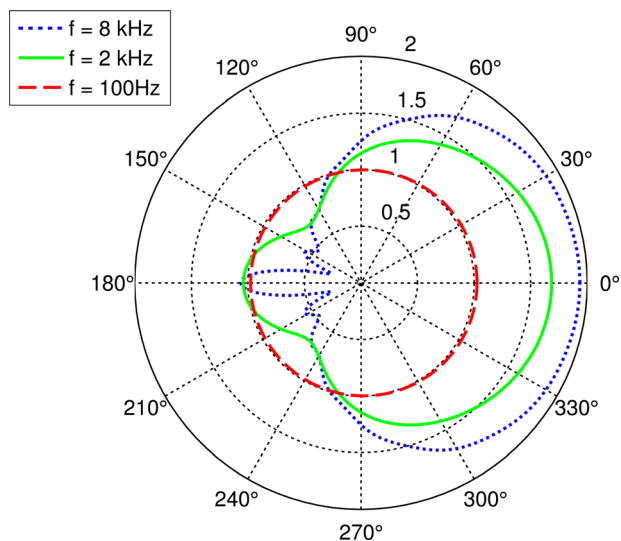

FIG. 3. (Color online) Polar plot of the magnitude of the response between a source and a receiver placed on a rigid sphere of radius 5 cm, at a distance of 1 m, for various frequencies.

6 dB magnitude gain compared to the open sphere at infinite frequency. The response on the back side of the rigid sphere is generally lower than in the open sphere case and lower than on the front side, as one would intuitively expect due to it being occluded. However at the very back of the sphere (i.e., the DOA is 180°) we observe a narrow *bright spot:* the waves propagating around the sphere all arrive in phase at the 180° point and as a result sum constructively.

A polar plot of the magnitude response (Fig. 3) illustrates both the near-doubling of the response on the front side of the sphere, and the bright spot on the back side of the sphere, which narrows as frequency increases. It should be noted that although the above-mentioned results are for a fixed sphere radius, as the scattering model is a function of kr, the effects of a change in radius are the same as a change in frequency; indeed the relevant factor is the radius of the sphere relative to the wavelength. These substantial differences between the open and rigid sphere responses confirm the need for a simulation method which accounts for scattering.

D. Proposed method

We now present our proposed method, incorporating the SHD presented in Sec. III A and the scattering model introduced in Sec. III B.

Due to the different formulation of the spherical harmonic domain Neumann Green's function compared to the spatial domain Green's function, as well as the directionality of the array's response, two changes must be made to the image position vectors $\mathbf{R_p}$ and $\mathbf{R_m}$ in our proposed method. First, to compute the SHD in the Neumann Green's function, we require the distance between each image and the *center* of the array [r_s in Eq. (11)]; this is accomplished by computing the image position vectors using the position of the center of the array rather than the position of the receiver. Second, to compute the SHD we require the angle of each image with respect to the center of the array [$\Theta_{\bar{r},\bar{r}_s}$ in Eq. (11)]. In Allen and Berkley's image method, the direction of the vector $\mathbf{R_p} + \mathbf{R_m}$ is not always the same: in some cases it points from the receiver to the image and in others it points from the image to the receiver. This is not an issue for the image method as only the norm of this vector is used. As we also require the angle of the images in our proposed method, we modify the definition of $\mathbf{R_p}$, such that the vector $\mathbf{R_p} + \mathbf{R_m}$ always points from the center of the array to the image.

We now incorporate these two changes into the definition of the image vectors $\mathbf{R_p}$ and $\mathbf{R_m}$. If the sound source is located at $\mathbf{r_s} = (x_s, y_s, z_s)$ and the center of the sphere is located at $\mathbf{r_a} = (x_a, y_a, z_a)$, the images obtained using the walls at $x = 0$, $y = 0$, and $z = 0$ are expressed as a vector $\mathbf{R_p}$,

$$\mathbf{R_p} = [x_s - 2p_x x_s - x_a, y_s - 2p_y y_s - y_a, z_s - 2p_z z_s - z_a]. \tag{15}$$

For brevity we define $\mathbf{R_{p,m}} \triangleq \mathbf{R_p} + \mathbf{R_m}$ allowing us to express the distance between an image and the center of the sphere as $\|\mathbf{R_{p,m}}\|$ and the angle of the image taken with respect to the center of the sphere as $\angle\mathbf{R_{p,m}}$. The image positions in the x dimension are illustrated in Fig. 4.
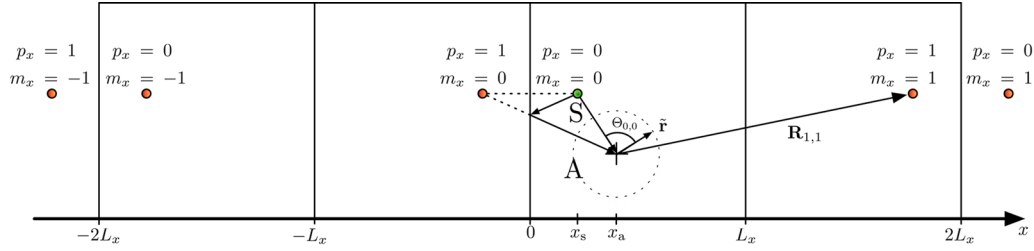
FIG. 4. (Color online) A slice through the image space showing the positions of the images in the x dimension, with a source S and array A. The full image space has three dimensions (x, y, and z). An example of a reflected path is shown for the image with $p_x = 1$ and $m_x = 0$.

If we denote the image positions in spherical coordinates as $\tilde{\mathbf{R}}_{\mathbf{p},\mathbf{m}}$, the RTF $H(\tilde{\mathbf{r}}|\tilde{\mathbf{r}}_s, k)$ is the weighted sum of the individual responses $G_N(\tilde{\mathbf{r}}|\tilde{\mathbf{R}}_{\mathbf{p},\mathbf{m}}, k)$ for each of the images[21]

$$H(\tilde{\mathbf{r}}|\tilde{\mathbf{r}}_s, k) = \sum_{\mathbf{p} \in \mathcal{P}} \sum_{\mathbf{m} \in \mathcal{M}} \beta_{x_1}^{|m_x - p_x|} \beta_{x_2}^{|m_x|} \beta_{y_1}^{|m_y - p_y|} \beta_{y_2}^{|m_y|}$$
$$\times \beta_{z_1}^{|m_z - p_z|} \beta_{z_2}^{|m_z|} G_N(\tilde{\mathbf{r}}|\tilde{\mathbf{R}}_{\mathbf{p},\mathbf{m}}, k). \quad (16)$$

As we are working in the wavenumber domain, we can allow for frequency-dependent boundary reflection coefficients in Eq. (16), if desired. The reflection coefficients would then be written as $\beta_{x_1}(k)$, $\beta_{x_2}(k)$ and so on. Chen and Maher[22] provide some measured reflection coefficients for a wall, window, floor and ceiling.

IV. IMPLEMENTATION

A. Truncation error

To compute the expression for the RTF in Eq. (16), the sum over an infinite number of orders l in the Neumann Green's function G_N must be approximated by a sum $\hat{G}_N$ over a finite order L. Choosing L too small will result in a large approximation error, whereas choosing L too large will result in too high a computational complexity. We now investigate the approximation error in order to provide some guidelines for the choice of the order L. The results for an open sphere are provided for reference, and were computed by using a truncated SHD of the Green's function $\hat{G}$ instead of a Neumann Green's function.

For an open sphere, the error can be determined exactly because the Green's function is a decomposition of the closed-form expression in Eq. (1). For a rigid sphere, however, no closed-form expression exists as the scattering term can be expressed only in the spherical harmonic domain. We therefore estimated the error by comparing the truncated Neumann Green's function $\hat{G}_N$ to a high-order Neumann Green's function. Based on simulations performed with an open sphere, where a true reference is available, we can safely assume that the error involved in using a high-order Neumann Green's function as a reference as opposed to the untruncated Neumann Green's function is small. In practice, we cannot choose very large values of L because of numerical difficulties involved in multiplying high order spherical Bessel and Hankel functions. For typical sphere radii and source–array distances, this allows us to reach L values of up to about 100 using our MATLAB implementation.[9]

We evaluated the truncated (Neumann) Green's function at $K = 1024$ discrete values of k (denoted by $\tilde{k}$), forming a set $\mathcal{K}$ corresponding to frequencies in the range 100 Hz–8 kHz,[23] and then calculated the normalized root-mean-square magnitude error ϵ_m and the root-mean-square phase error ϵ_p, i.e.,

$$\epsilon_m(\tilde{\mathbf{r}}|\tilde{\mathbf{r}}_s, L) = \sqrt{\frac{1}{K} \sum_{\tilde{k} \in \mathcal{K}} \frac{\left(|G_N(\tilde{\mathbf{r}}|\tilde{\mathbf{r}}_s, \tilde{k})| - |\hat{G}_N(\tilde{\mathbf{r}}|\tilde{\mathbf{r}}_s, \tilde{k}, L)|\right)^2}{|G_N(\tilde{\mathbf{r}}|\tilde{\mathbf{r}}_s, \tilde{k})|^2}},$$
(17)

$$\epsilon_p(\tilde{\mathbf{r}}|\tilde{\mathbf{r}}_s, L) = \sqrt{\frac{1}{K} \sum_{\tilde{k} \in \mathcal{K}} \left(\angle G_N(\tilde{\mathbf{r}}|\tilde{\mathbf{r}}_s, \tilde{k}) - \angle \hat{G}_N(\tilde{\mathbf{r}}|\tilde{\mathbf{r}}_s, \tilde{k}, L)\right)^2}.$$
(18)

We averaged the magnitude and phase errors over 32 quasi-equidistant receivers and 50 random source positions at a fixed distance from the center of the array.

The resulting average errors are given in Fig. 5, for both the open and rigid sphere cases. Three different sphere radii were used: $r = 4.2$ cm (the radius of the Eigenmike[24]), $r = 10$ cm and $r = 15$ cm. A source–array distance of 1 m was used; results for 1–5 m are omitted as they are essentially identical. It can be seen that beyond a certain threshold, increases in L give only a very small reduction in error. A
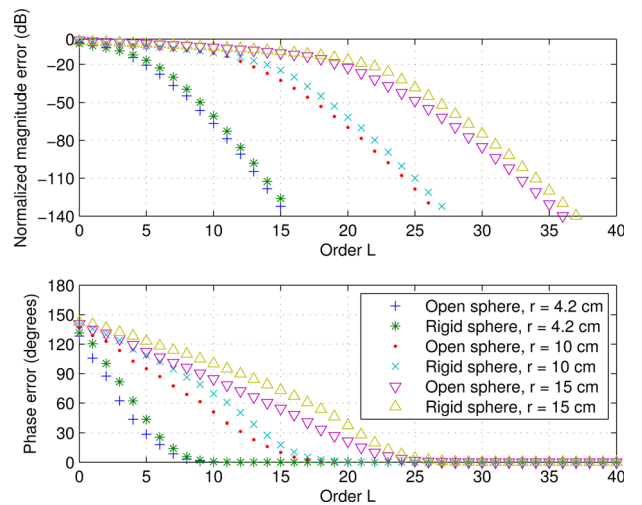

FIG. 5. (Color online) Magnitude and phase errors involved in the truncation of the SHD in the Green's function (open sphere) and the Neumann Green's function (rigid sphere).

Jarrett et al.: Rigid sphere room impulse response simulation

rule of thumb for choosing L is $L > \lceil 1.1 k_{\max} r \rceil$, where $k_{\max}$ is the largest wavenumber of interest.

B. Computational complexity

As the RTFs are made up of a sum over all orders l that includes spherical Hankel functions h_l and Legendre polynomials P_l, we can make use of recursion relations over l to reduce the computational complexity of these functions. For the spherical Hankel function, we make use of the following relation:[25]

$$h_m(x) = \frac{2m-1}{x} h_{m-1}(x) - h_{m-2}(x), \quad m \geq 2, \qquad (19)$$

where

$$h_0(x) = \frac{e^{ix}}{ix}, \quad h_1(x) = \frac{e^{ix}}{ix^2} - \frac{e^{ix}}{x}. \qquad (20)$$

For the Legendre polynomials we use a similar recursion relation[25]

$$P_m(x) = \frac{2m-1}{m} x P_{m-1}(x) - \frac{m-1}{m} P_{m-2}(x), \quad m \geq 2, \qquad (21)$$

where $P_0(x) = 1$ and $P_1(x) = x$.

While replacing the exponential in Eq. (1) with a SHD does lead to an increase in computational complexity when computing the RTF for a single receiver (which is unavoidable in the rigid sphere case), this can have an advantage when simulating many receiver positions. For the conventional image method, we must compute the image positions and resulting response separately for each individual receiver. However, in the proposed method the image positions are all computed with respect to the *center* of our array, and therefore only once for all of the microphones in the array.

An alternative to Eq. (16) is obtained by changing the order of the summations in the RTF and computing the sum over all images only once, instead of once per receiver, i.e.,

$$\begin{aligned}
H(\tilde{\mathbf{r}}|\tilde{\mathbf{r}}_s, k) = ik \sum_{l=0}^{\infty} \sum_{m=-l}^{l} Y_{lm}(\Omega) \sum_{\mathbf{p} \in \mathcal{P}} \sum_{\mathbf{m} \in \mathcal{M}} \beta_{x_1}^{|m_x - p_x|} \\
\times \beta_{x_2}^{|m_x|} \beta_{y_1}^{|m_y - p_y|} \beta_{y_2}^{|m_y|} \beta_{z_1}^{|m_z - p_z|} \beta_{z_2}^{|m_z|} \\
\times b_l(kr) h_l^{(1)}(k\|\mathbf{R}_{\mathbf{p},\mathbf{m}}\|) Y_{lm}^*(\angle \mathbf{R}_{\mathbf{p},\mathbf{m}}). \qquad (22)
\end{aligned}$$

The expression in Eq. (22) requires $O((N+M)(L+1)^2)$ operations per discrete frequency, where L is the maximum SHD order, N is the number of images and M is the number of microphones, whereas the approach in Eq. (16) requires $O(N M(L+1))$ operations per discrete frequency. As the number of images N is typically very large, $(N+M)(L+1)^2 \approx N(L+1)^2$. Assuming the operations in the two approaches are of similar complexity, it is therefore more efficient to use the expression in Eq. (16) for $M < L+1$ and the expression in Eq. (22) for $M > L+1$. Consequently the least computa-

tionally complex approach depends on the number of microphones M and array radius r. In the remainder of this paper we use the expression in Eq. (16); this is particularly appropriate in the applications in Sec. V B where $M = 2$ and in Sec. V C where $M = 1$.

C. Algorithm summary

A summary of the proposed method is presented in the form of pseudocode in Fig. 6. The variable *nsample* denotes the number of samples in the RIR and N_o the maximum reflection order.

The number of computations has been reduced by processing only half of the frequency spectrum because we know the RIR is real and the corresponding RTF is conjugate

1: $\mathcal{P} = \{0,1\}^3$
2: $\mathcal{M} = \{-N_m, \cdots, 0, \cdots, N_m\}^3$
3: $\mathcal{A} = \mathcal{P} \times \mathcal{M}$

4: **for** $(\mathbf{p}, \mathbf{m}) \in \mathcal{A}$ **do**
5:    **if** $|2m_x - p_x| + |2m_y - p_y| + |2m_z - p_z| \leq N_o$ **then**
6:      $\mathbf{R}_{\mathbf{p},\mathbf{m}} = \begin{bmatrix} x_s - 2p_x x_s - x_a + 2m_x L_x \\ y_s - 2p_y y_s - y_a + 2m_y L_y \\ z_s - 2p_z z_s - z_a + 2m_z L_z \end{bmatrix}$
7:      $\beta(\mathbf{p}, \mathbf{m}) = \beta_{x_1}^{|m_x - p_x|} \beta_{x_2}^{|m_x|} \beta_{y_1}^{|m_y - p_y|} \beta_{y_2}^{|m_y|} \beta_{z_1}^{|m_z - p_z|} \beta_{z_2}^{|m_z|}$
8:    **else**
9:      $\mathcal{A} = \mathcal{A} \setminus \{(\mathbf{p}, \mathbf{m})\}$
10:    **end if**
11: **end for**

12: **for** $k = 1 \to nsample/2 + 1$ **do**
13:    **for** $l = 0 \to L$ **do**
14:      **if** $sphType =$ 'rigid' **then**
15:        $\Delta(k,l) = \frac{i}{h_l^{(1)'}(kr)(kr)^2}$
16:      **else**
17:        $\Delta(k,l) = j_l(kr)$
18:      **end if**
19:      $\Gamma(k,l) = \frac{i \cdot k}{4\pi} \cdot \Delta(k,l)$
20:    **end for**
21: **end for**

22: **for** $(\mathbf{p}, \mathbf{m}) \in \mathcal{A}$ **do**
23:    **if** $\|\mathbf{R}_{\mathbf{p},\mathbf{m}}\| + r < c \cdot nsample/fs$ **then**
24:      **for** $ang = 1 \to M$ **do**
25:        $\Theta = \hat{\mathbf{R}}_{\mathbf{p},\mathbf{m}} \cdot \hat{\mathbf{r}}(ang)$
26:        **for** $l = 0 \to L$ **do**
27:          $\Upsilon(ang,l) = P_l(\Theta) \cdot (2l+1)$
28:        **end for**
29:      **end for**
30:      **for** $k = 1 \to nsample/2 + 1$ **do**
31:        **for** $l = 0 \to L$ **do**
32:          $\Lambda(k,l) = h_l(k\|\mathbf{R}_{\mathbf{p},\mathbf{m}}\|) \cdot \Gamma(k,l)$
33:        **end for**
34:      **end for**
35:      **for** $ang = 1 \to M$ **do**
36:        **for** $k = 1 \to nsample/2 + 1$ **do**
37:          **for** $l = 0 \to L$ **do**
38:            $H(\mathbf{p}, \mathbf{m}, ang, k, l) = \beta(\mathbf{p}, \mathbf{m}) \cdot \Upsilon(ang,l) \cdot \Lambda(k,l)$
39:          **end for**
40:          $H(\mathbf{p}, \mathbf{m}, ang, k) = \sum_l H(\mathbf{p}, \mathbf{m}, ang, k, l)$
41:        **end for**
42:      **end for**
43:    **end if**
44: **end for**
45: $H(ang, k) = \sum_{(\mathbf{p},\mathbf{m}) \in \mathcal{A}} H(\mathbf{p}, \mathbf{m}, ang, k)$
46: $h(ang, n) = \text{IFFT}_R\{H(ang, k)\}$

FIG. 6. Pseudocode for the proposed method.

symmetric. The pseudocode necessary to compute the Hankel functions and Legendre polynomials is omitted here, as their computation is straightforward using recursion relations (19) and (21).

SMIRGEN, a MATLAB/C++ implementation of the method in the form of a MEX-function is also available online.[9]

V. EXAMPLES & APPLICATIONS

In this section we give a number of examples that make use of the proposed method. Wherever possible we compared the simulated results to theoretical results obtained using approximate models. These examples are given to illustrate and partially validate the proposed method.

A. Diffuse sound field energy

In statistical room acoustics (SRA), reverberant sound fields are modeled as diffuse sound fields, allowing for a statistical analysis of reverberation instead of computing each of the individual reflections. In this section, we compare a theoretical prediction of sound energy on the surface of a rigid sphere, based on a diffuse model of reverberation, to simulated results obtained using the proposed method.

A diffuse sound field is composed of mutually independent plane waves incident from all directions with equal probability and amplitude.[26] Using the scattering model previously introduced, we can determine the cross-correlation between the sound pressure at positions $\tilde{\mathbf{r}}$ and $\tilde{\mathbf{r}}'$ on the surface of a sphere, due to a diffuse sound field (see the Appendix for derivation)[27]

$$C(\tilde{\mathbf{r}}, \tilde{\mathbf{r}}', k) = \sum_{l=0}^{\infty} |b_l(kr)|^2 (2l+1) P_l(\cos \Theta_{\tilde{\mathbf{r}}, \tilde{\mathbf{r}}'}), \qquad (23)$$

where $\Theta_{\tilde{\mathbf{r}}, \tilde{\mathbf{r}}'}$ is the angle between $\tilde{\mathbf{r}}$ and $\tilde{\mathbf{r}}'$. In the open sphere case, it is shown in the Appendix that this simplifies to the well-known spatial domain expression[26,28,29] $\text{sinc}(k\|\tilde{\mathbf{r}} - \tilde{\mathbf{r}}'\|)$, where sinc denotes the unnormalized sinc function.

For the sound energy at a position $\tilde{\mathbf{r}}$ we substitute $\Theta_{\tilde{\mathbf{r}}, \tilde{\mathbf{r}}'} = 0$ and find $C(\tilde{\mathbf{r}} - \tilde{\mathbf{r}}', k) = \sum_{l=0}^{\infty} |b_l(kr)|^2 (2l+1)$. According to SRA theory, for frequencies above the Schroeder frequency[26] the energy of the reverberant sound field H_r is then given by[28]

$$E\{|H_r(\tilde{\mathbf{r}}, k)|^2\} = \frac{1-\alpha}{\pi A \alpha} C(\tilde{\mathbf{r}}, \tilde{\mathbf{r}}, k)$$
$$= \frac{1-\alpha}{\pi A \alpha} \sum_{l=0}^{\infty} |b_l(kr)|^2 (2l+1), \qquad (24)$$

where $E\{\cdot\}$ denotes spatial expectation, α is the average wall absorption coefficient and A is the total wall surface area.

The above-presented theoretical expression for the average reverberant energy can be compared to simulated results obtained using our method. We computed the spatial expectation using an average over 200 source–array positions, using the approach in Radlović et al.:[29] the array and source
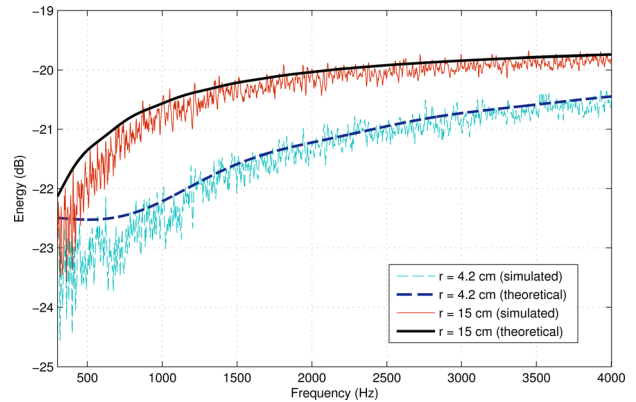

FIG. 7. (Color online) Theoretical and simulated reverberant sound field energy on the surface of a rigid sphere, as a function of frequency for two array radii. The simulated results are averaged over 200 source–array positions, all at least half a wavelength from the room boundaries.

were kept in a fixed configuration (at a distance of 2 m from each other), which was then randomly rotated and translated. Both sources and microphones were kept at least half a wavelength from the boundaries of the room, helping to ensure the diffuseness of the reverberant sound field.[26] The reverberant component H_r of the RTFs was computed by subtracting the direct path H_d from the simulated RTFs.

The room dimensions were equal to $6.4 \times 5 \times 4$ m, as in Radlović et al.,[29] in order to best approximate a diffuse sound field. The reverberation time RT_{60} was set to 500 ms, giving an average wall absorption coefficient of $\alpha = 0.2656$. We simulated RIRs with a length of 4096 samples at a sampling frequency of 8 kHz. We considered frequencies from 300 Hz to 4 kHz, well above the Schroeder frequency of $2000\sqrt{0.5/(4 \cdot 5 \cdot 6.4)} = 125$ Hz, and the half-wavelength minimum distance is therefore 57 cm for a speed of sound of 343 m/s. We averaged the results over the 200 source–array positions and 32 quasi-equidistant receiver positions.

In Fig. 7, we plot the theoretical and simulated energy of H_r as a function of frequency, for two array radii (4.2 and 15 cm). We note that, except at low frequencies, there is a good match between the theoretical diffuse field energy expression we derived and the results obtained using our method. At lower frequencies, the theoretical equation overestimates the energy; we hypothesize that this is due to the reverberant sound field still not being fully diffuse.

B. Binaural interaural time and level differences

The topic of binaural sound and in particular head-related transfer functions (HRTFs) or head-related impulse responses (HRIRs) is of interest to researchers and engineers working on surround sound reproduction, who for example aim to reproduce spatial audio through a pair of stereo headphones. In addition, the psychoacoustic community is interested in the ability of the human brain to localize sound sources using only two ears.

Two binaural cues that contribute to sound source localization in humans are the interaural time difference (ITD) and the interaural level (or intensity) difference (ILD).[30] The

Jarrett et al.: Rigid sphere room impulse response simulation

ITD measures the difference in arrival time of a sound at the two ears, and the ILD measures the level difference between the two ears. In this example, we study the long-term cues assuming the source signal is spectrally white. Therefore, we can compute the cues directly using the simulated RTFs.

We used the proposed method to simulate a simple HRTF by considering microphones placed at locations on a rigid sphere corresponding to ear positions on the human head. Although real HRTFs vary from individual to individual, depending on the head, torso, and pinnae, the main characteristics of the HRTF are also exhibited by a simple rigid sphere RTF.[19] The representation of HRTFs using spherical harmonics was studied in Refs. [31] and [32].

Whereas HRTFs do not include the effects of reverberation, and as a result typically sound artificial and provide poor cues for the perception of sound source distance,[33] the proposed method also allows for the inclusion of reverberation in HRIRs. In this case, they are then referred to as binaural room impulse responses (BRIRs). BRIRs are important for the analysis of the effects of reverberation on auditory perception, for example its impact on localization accuracy. As rotational symmetry no longer necessarily holds once the room reflections are taken into account, the measurement of BRIRs must be done for every source–head position and orientation and is therefore very time-consuming. Simulating BRIRs allows us to more easily study the effects of early and late reflections on the binaural cues.

We begin by looking at ITDs in an anechoic environment, in order to illustrate the effect of the head in isolation. We compare simulated results to approximate theoretical results provided by a ray-tracing formula attributed to Woodworth and Schlosberg that looks at the distance traveled from the source to an observation point on the sphere, either in free-space if the observation point is on the near side of the sphere, or via a point of tangency if the observation point is on the far side.[19]

The simulated results were obtained by generating HRIRs at a sampling frequency of 32 kHz, with a sphere ra-

dius of 8.75 cm and microphones placed at (0°, 100°) (corresponding to the left ear) and (0°, 260°) (corresponding to the right ear). The HRIRs were then band-pass filtered between 2.8 and 3.2 kHz.[34] The DOA was varied by rotating the source around the sphere at a fixed distance of 1 m and elevation of 0°. The simulated ITD was computed by determining the time delay that maximized the interaural cross-correlation between the two simulated and band-pass filtered HRIRs. The cross-correlation was interpolated using a second-order polynomial in order to obtain sub-sample delays.

In Fig. 8 we plot the ITD as a function of direction of arrival, where 0° corresponds to the median plane on the front side of the sphere and 180° corresponds to the median plane on the back side of the sphere. As expected, as the DOA increases from 0° to 80° and the source gets closer to the ipsilateral ear, the ITD increases monotonically until it reaches its maximum at 80°, at which point the source is furthest from the contralateral ear. The ITD then decreases from 80° to 180° as the source nears the median plane and gets closer to the contralateral ear. The response from 180° to 360° is not shown due to the symmetry about 180°. As we expect, our simulated results are reasonably close to the theoretical ray-tracing results,[19] with a difference of less than 70 μs.

Using the proposed method, we analyzed the ILDs in a reverberant environment under three scenarios: the sphere was either placed in the center of the room with a DOA of 0° (where the source is equidistant from the two ears), or at a distance of ∼0.5 m from one of the walls with DOAs of 0° and 100° (where the source is aligned with the left ear). In all three cases the source was placed at a distance of 1 m from the center of the sphere. We chose a room size of $9 \times 5 \times 3$ m with a reverberation time RT_{60} of 500 ms, and simulated BRIRs with a length of 4096 samples at a sampling frequency of 8 kHz.

In Figs. 9–11 we plot the ILDs for the three above-mentioned cases, as well as the ILDs we would obtain in an anechoic environment, which are entirely due to scattering.
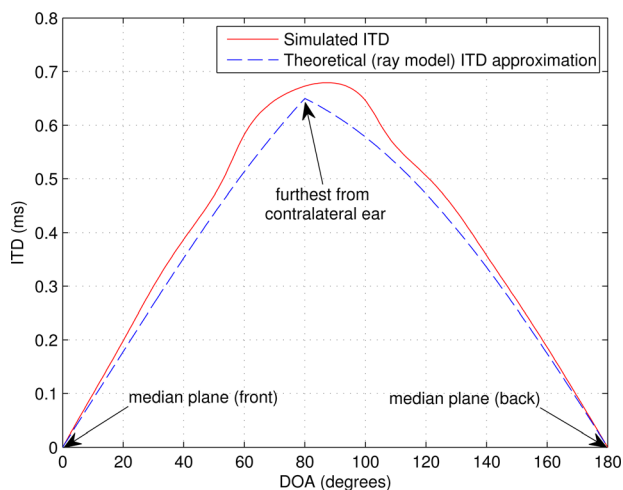

FIG. 8. (Color online) Comparison of ITDs as a function of source DOA, in simulation and using the theoretical ray model approximation. The simulated ITDs are based on HRIRs computed using our proposed algorithm in an anechoic environment.
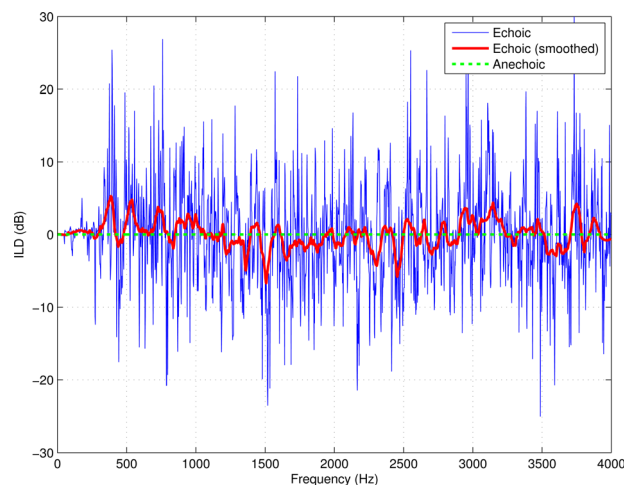

FIG. 9. (Color online) Comparison of ILDs in echoic and anechoic environments, with the sphere placed in the center of the room and a DOA of 0°. The ILDs are based on HRTFs (anechoic) and BRIRs (echoic) computed using the proposed method.
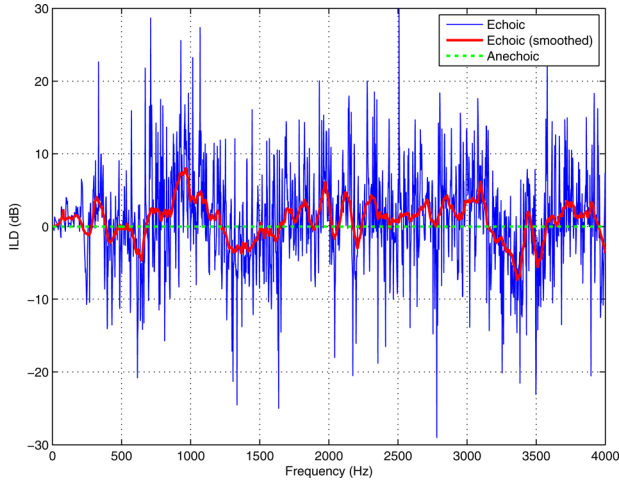
J. Acoust. Soc. Am., Vol. 132, No. 3, September 2012

Jarrett *et al.*: Rigid sphere room impulse response simulation   1469

FIG. 10. (Color online) Comparison of ILDs in echoic and anechoic environments, with the sphere placed near a room wall and a DOA of 0°.

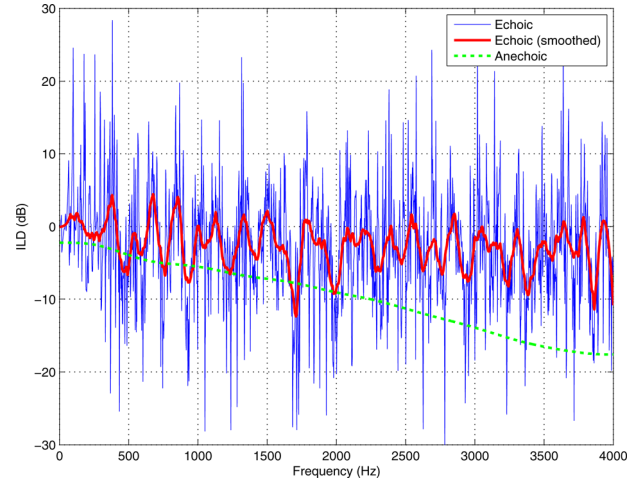FIG. 11. (Color online) Comparison of ILDs in echoic and anechoic environments, with the sphere placed near a room wall and a DOA of 100°.

The ILDs were computed by taking the difference in magnitude between the left ear response and the right ear response. A negative ILD therefore indicates that the magnitude of the ipsilateral ear response is lower than that of the contralateral ear response. The smoothed echoic ILDs were obtained using a Savitzky–Golay smoothing filter.

The main effect of reverberation we can observe is the introduction of random frequency-to-frequency variations; these are particularly obvious when most of the reverberant energy is diffuse, i.e., when the sphere is placed in the center of the room (Fig. 9). Room reflections also increase the overall reverberant energy, particularly in the contralateral ear, which receives less direct path energy, thus reducing the ILDs. This is especially noticeable when the contralateral ear is placed near a wall: the contralateral ear receives more energy than in the anechoic case and the ILD is therefore closer to zero (Fig. 11).

Placement of the sphere near a wall additionally introduces systematic distortions in the ILDs associated with the prominent early reflection from this wall. The resonant frequencies of the room (or eigenfrequencies)[26] are dominated by those associated with this near wall; this is visible in Fig. 11 and most noticeably in Fig. 10.

All these effects have also been observed experimentally with a manikin by Shinn-Cunningham et al.[33] The proposed algorithm is therefore a good way of predicting the effects of head movement and environmental changes (such as reverberation time) on HRTFs or BRIRs, without the need for more cumbersome experiments with head and torso simulators for example.

C. Mouth simulator

The principle of reciprocity can often be advantageously used in room acoustics measurements. The principle states that RTFs are symmetric in the coordinates of the sound source and the observation point: "If we put the sound source at **r**, we observe at point **r**$_0$ the same sound pressure as we did before at **r**, when the sound source was at **r**$_0$."[26] We can apply this principle to RTF simulations, and use our method

to generate the RTF between one or more sources on a sphere and a single omnidirectional microphone placed away from the sphere.

A specific application of this is a mouth simulator: we model the head as a rigid sphere (as in Sec. V B) of radius r_h, and the mouth as an omnidirectional point source placed on this rigid sphere. This is straightforward implemented in the proposed method by replacing the source position with the microphone position $\tilde{\mathbf{r}}_{mic}$, the microphone position with the mouth position $\tilde{\mathbf{r}}_{mouth} = (r_h, \Omega_{mouth})$, and the array position with the head position, i.e.,

$$H(\tilde{\mathbf{r}}_{mic}|\tilde{\mathbf{r}}_{mouth}, k) = H(\tilde{\mathbf{r}} = \tilde{\mathbf{r}}_{mouth}|\tilde{\mathbf{r}}_s = \tilde{\mathbf{r}}_{mic}, k).$$

As a result we can simulate the RTF between a mouth on a head, and a single microphone in free space. Repeated use of the algorithm allows for multiple receivers.

Although more accurate modeling of the head and mouth is possible using finite element or boundary element methods, for example, our algorithm is valuable for application to this problem due its comparative simplicity and the fact that, if desired, it can also take into account room reverberation. While the diameter of the mouth plays an important role in determining the filter characteristic of the vocal tract,[37] we assume for the purposes of the scattering model that the mouth is a point source.

As an illustration of this application, Fig. 12 shows the energy of the RTF between the mouth and a microphone as a function of microphone position at frequencies of 100 Hz and 3 kHz in an anechoic environment. The mouth was positioned on a sphere of radius 8.75 cm. Only two dimensions, x and y, are shown for brevity as the z dimension is identical to x and y. We observe that at 100 Hz there is no scattering and the radiation pattern is omnidirectional so that the sphere has little effect. At 3 kHz the effect of scattering starts to become more significant, and the energy at the back of the sphere is reduced, whereas the energy at the front is increased. Finally the bright spot discussed in Sec. III C is particularly apparent at the very back of the sphere in Fig. 12(b).
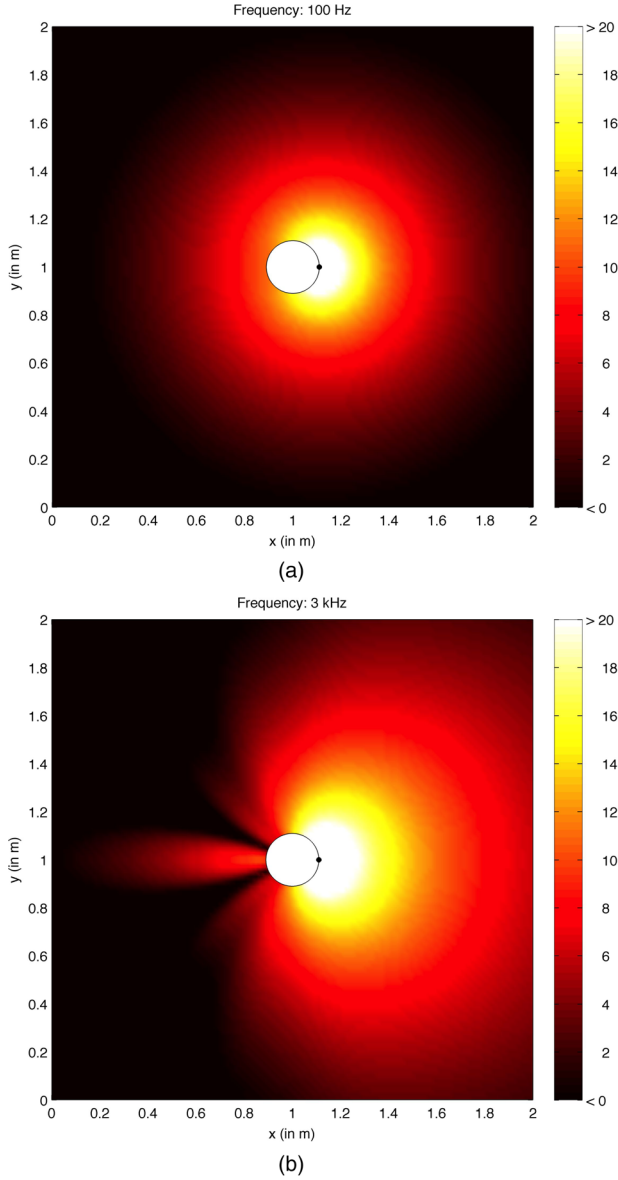
FIG. 12. (Color online) Sound energy radiation pattern (dB) at 100 Hz (a) and 3 kHz (b). The mouth position is denoted by a black dot.

VI. CONCLUSION

Spherical microphone arrays on a rigid baffle are of great interest currently. In order to analyze, work with, and develop acoustic signal processing algorithms that make use of a spherical microphone array, a simulator is desired that can take into account the effects of the acoustic environment of the array, as well as the scattering effects of the rigid spherical baffle. Accordingly a method was presented for the simulation of RIRs or RTFs for a rigid spherical microphone array in a reverberant environment.

We presented a scattering model used to model the rigid sphere, justifying its use with references to the literature, and provided an overview of the model's behavior. We showed that the error with respect to the theoretical model can be controlled at the expense of increased computational complexity. Finally we provided a number of examples showing additional applications of this method.

APPENDIX: SPATIAL CORRELATION IN A DIFFUSE SOUND FIELD

The sound pressure at a position $\tilde{\mathbf{r}} = (r, \Omega)$ due to a unit amplitude plane wave incident from direction Ω_0 is given by[16]

$$p(\tilde{\mathbf{r}}, \Omega_0, k) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} 4\pi \varphi(\Omega_0)(-i)^l b_l(kr) Y_{lm}^*(\Omega_0) Y_{lm}(\Omega),$$
(A1)

where $\varphi(\Omega_0)$ is a random phase term and $|\varphi(\Omega_0)| = 1$. Assuming a diffuse sound field, the spatial cross-correlation between the sound pressure at two positions $\tilde{\mathbf{r}} = (r, \Omega)$ and $\tilde{\mathbf{r}}' = (r, \Omega')$ is given by

$$C(\tilde{\mathbf{r}}, \tilde{\mathbf{r}}', k) = \frac{1}{4\pi} \int_{\Omega_0 \in S^2} p(\tilde{\mathbf{r}}, \Omega_0, k) p^*(\tilde{\mathbf{r}}', \Omega_0, k) d\Omega_0$$

$$= \frac{1}{4\pi} \int_{\Omega_0 \in S^2} \sum_{l=0}^{\infty} \sum_{m=-l}^{l} 4\pi(-i)^l b_l(kr) Y_{lm}^*(\Omega_0) Y_{lm}(\Omega)$$

$$\times \sum_{l'=0}^{\infty} \sum_{m'=-l'}^{l'} 4\pi i^{l'} b_{l'}^*(kr) Y_{l'm'}(\Omega_0) Y_{l'm'}^*(\Omega') d\Omega_0.$$

Using the orthonormality property of the spherical harmonics[16] $\int_{\Omega \in S^2} Y_{lm}(\Omega) Y_{l'm'}^*(\Omega) d\Omega = \delta_{ll'} \delta_{mm'}$ and the addition theorem in Eq. (8), we eliminate the cross terms followed by the sum over m and obtain

$$C(\tilde{\mathbf{r}}, \tilde{\mathbf{r}}', k) = \frac{1}{4\pi} \sum_{l=0}^{\infty} \sum_{m=-l}^{l} (4\pi)^2 |b_l(kr)|^2 Y_{lm}(\Omega) Y_{lm}^*(\Omega'),$$

$$= \frac{1}{4\pi} \sum_{l=0}^{\infty} (4\pi)^2 |b_l(kr)|^2 \frac{2l+1}{4\pi} P_l(\cos \Theta_{\tilde{\mathbf{r}}, \tilde{\mathbf{r}}'})$$

$$= \sum_{l=0}^{\infty} |b_l(kr)|^2 (2l+1) P_l(\cos \Theta_{\tilde{\mathbf{r}}, \tilde{\mathbf{r}}'}), \quad \text{(A2)}$$

where $\Theta_{\tilde{\mathbf{r}}, \tilde{\mathbf{r}}'}$ is the angle between $\tilde{\mathbf{r}}$ and $\tilde{\mathbf{r}}'$.

In the open sphere case where $b_l(kr) = j_l(kr)$, we can express Eq. (A2) as

$$C(\tilde{\mathbf{r}}, \tilde{\mathbf{r}}', k) = \Im \left\{ 4\pi i \sum_{l=0}^{\infty} \sum_{m=-l}^{l} |b_l(kr)|^2 Y_{lm}(\Omega) Y_{lm}^*(\Omega') \right\}$$

$$= \Im \left\{ 4\pi i \sum_{l=0}^{\infty} \sum_{m=-l}^{l} j_l(kr) h_l^{(1)}(kr) Y_{lm}(\Omega) Y_{lm}^*(\Omega') \right\}$$

using $\Re\{h_l^{(1)}(kr)\} = j_l(kr)$, where $\Re$ and $\Im$, respectively, denote the real and imaginary parts of a complex number. Finally, using Eq. (7), we obtain the well-known spatial domain result for two omnidirectional receivers in a diffuse sound field[26,28,29]

$$C(\tilde{\mathbf{r}}, \tilde{\mathbf{r}}', k) = \Im \left\{ \frac{e^{+ik\|\tilde{\mathbf{r}} - \tilde{\mathbf{r}}'\|}}{k\|\tilde{\mathbf{r}} - \tilde{\mathbf{r}}'\|} \right\}$$

$$= \frac{\sin(k\|\tilde{\mathbf{r}} - \tilde{\mathbf{r}}'\|)}{k\|\tilde{\mathbf{r}} - \tilde{\mathbf{r}}'\|}. \quad \text{(A3)}$$

[1]J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," J. Acoust. Soc. Am. **65**, 943–950 (1979).

[2]P. M. Peterson, "Simulating the response of multiple microphones to a single acoustic source in a reverberant room," J. Acoust. Soc. Am. **80**, 1527–1529 (1986).

[3]E. Lehmann and A. Johansson, "Diffuse reverberation model for efficient image-source simulation of room impulse responses," IEEE Trans. Audio, Speech, Lang. Process. **18**, 1429–1439 (2010).

[4]T. D. Abhayapala and D. B. Ward, "Theory and design of high order sound field microphones using spherical microphone array," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2002), Vol. 2, pp. 1949–1952.

[5]J. Meyer and G. Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2002), Vol. 2, pp. 1781–1784.

[6]B. Rafaely, "Plane-wave decomposition of the pressure on a sphere by spherical convolution," J. Acoust. Soc. Am. **116**, 2149–2157 (2004).

[7]N. Gumerov and R. Duraiswami, "Modeling the effect of a nearby boundary on the HRTF," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2001), Vol. 5, pp. 3337–3340.

[8]T. Betlehem and M. A. Poletti, "Sound field reproduction around a scatterer in reverberation," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2009), pp. 89–92.

[9]D. P. Jarrett, "Spherical microphone array impulse response (SMIR) generator," http://www.ee.ic.ac.uk/sap/smirgen/.

[10]D. P. Jarrett, E. A. P. Habets, M. R. P. Thomas, and P. A. Naylor, "Simulating room impulse responses for spherical microphone arrays," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2011), pp. 129–132.

[11]This expression assumes the sign convention commonly used in physics/acoustics, whereby the temporal Fourier transform is defined as $F(\omega) = \int_{-\infty}^{\infty} f(t)e^{+i\omega t}dt$ in order to eliminate the $e^{-i\omega t}$ term in the time-harmonic solution to the wave equation, as in Refs. 14 and 16. The formulas in this paper are the complex conjugates of those found in other papers which use the opposite sign convention.

[12]Some texts (Ref. 19) refer to the scattering effect as diffraction, although Morse and Ingard note that "When the scattering object is large compared with the wavelength of the scattered sound, we usually say the sound is reflected and diffracted, rather than scattered" (Ref. 14), therefore, in the case of spherical microphone arrays (particularly rigid ones, which tend to be relatively small for practical reasons), scattering is possibly the more appropriate term.

[13]N. A. Logan, "Survey of some early studies of the scattering of plane waves by a sphere," Proc. IEEE **53**, 773–785 (1965).

[14]P. M. Morse and K. U. Ingard, *Theoretical Acoustics, International Series in Pure and Applied Physics* (McGraw-Hill, New York, 1968), p. 927.

[15]D. L. Sengupta, "The sphere," in *Electromagnetic and Acoustic Scattering by Simple Shapes*, edited by J. J. Bowman, T. B. A. Senior, and P. L. E. Uslenghi (North-Holland, Amsterdam, The Netherlands, 1969), Chap. 10, pp. 353–415.

[16]E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*, 1st ed. (Academic, London, 1999), p. 306.

[17]J. Meyer and G. W. Elko, "Position independent close-talking microphone," Signal Process. **86**, 1254–1259 (2006).

[18]B. Rafaely, "Analysis and design of spherical microphone arrays," IEEE Trans. Speech Audio Process. **13**, 135–143 (2005).

[19]R. O. Duda and W. L. Martens, "Range dependence of the response of a spherical head model," J. Acoust. Soc. Am. **104**, 3048–3058 (1998).

[20]Z. Li and R. Duraiswami, "Hemispherical microphone arrays for sound capture and beamforming," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* (2005), pp. 106–109.

[21]The sign in the powers of β is different from that in Allen and Berkley's conventional image method, due to the change in the definition of $\mathbf{R}_p$.

[22]Z. Chen and R. C. Maher, "Addressing the discrepancy between measured and modeled impulse responses for small rooms," in *Proceedings of the Audio Engineering Society Conventions* (2007), pp. 1–14.

[23]Very low frequencies are omitted due to the fact that the spherical Hankel function $h_l(x)$ has a singularity around $x = 0$.

[24]J. Meyer and T. Agnello, "Spherical microphone array for spatial sound recording," in *Proceedings of the Audio Engineering Society Conventions*, New York (2003), pp. 1–9.

[25]*Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, edited by M. Abramowitz and I. A. Stegun (Dover Publications, New York, 1972), p. 1046.

[26]H. Kuttruff, *Room Acoustics*, 4th ed. (Taylor & Francis, London, 2000), p. 349.

[27]D. P. Jarrett, E. A. P. Habets, M. R. P. Thomas, N. D. Gaubitch, and P. A. Naylor, "Dereverberation performance of rigid and open spherical microphone arrays: Theory & simulation," in *Proceedings of the Joint Workshop on Hands-Free Speech Communication and Microphone Arrays (HSCMA)*, Edinburgh, UK (2011), pp. 145–150.

[28]D. B. Ward, "On the performance of acoustic crosstalk cancellation in a reverberant environment," J. Acoust. Soc. Am. **110**, 1195–1198 (2001).

[29]B. D. Radlović, R. Williamson, and R. Kennedy, "Equalization in an acoustic reverberant environment: Robustness results," IEEE Trans. Speech Audio Process. **8**, 311–319 (2000).

[30]T. T. Sandel, D. C. Teas, W. E. Feddersen, and L. A. Jeffress, "Localization of sound from single and paired sources," J. Acoust. Soc. Am. **27**, 842–852 (1955).

[31]A. Avni and B. Rafaely, "Sound localization in a sound field represented by spherical harmonics," in *Proceedings of the Second International Symposium on Ambisonics and Spherical Acoustics*, Paris, France (2010), pp. 1–5.

[32]M. J. Evans, J. A. S. Angus, and A. I. Tew, "Analyzing head-related transfer function measurements using surface spherical harmonics," J. Acoust. Soc. Am. **104**, 2400–2411 (1998).

[33]B. G. Shinn-Cunningham, N. Kopco, and T. J. Martin, "Localizing nearby sound sources in a classroom: Binaural room impulse responses," J. Acoust. Soc. Am. **117**, 3100–3115 (2005).

[34]Although the ray-tracing formula is frequency independent, it has been shown (Ref. 35) that ITDs actually exhibit some frequency dependence, and that because the ray-tracing concept applies to short wavelengths, this model yields only the high frequency time delay. Kuhn provides a more comprehensive discussion of this model and the frequency dependence of ITDs (Ref. 36). It should be noted the simulation results in Fig. 8 are in broad agreement with Kuhn's measured results at 3.0 kHz.

[35]C. Brown and R. Duda, "A structural model for binaural sound synthesis," IEEE Trans. Speech Audio Process. **6**, 476–488 (1998).

[36]G. F. Kuhn, "Model for the interaural time differences in the azimuthal plane," J. Acoust. Soc. Am. **62**, 157–167 (1977).

[37]J. R. Deller, J. G. Proakis, and J. H. L. Hansen, *Discrete-Time Processing of Speech Signals* (MacMillan, New York, 1993), p. 928.