SIMULATING ROOM IMPULSE RESPONSES FOR SPHERICAL MICROPHONE ARRAYS

Daniel P. Jarrett, Emanuël A. P. Thomas, Mark R. P. Thomas, Patrick A. Naylor

Dept. of Electrical & Electronic Engineering, Imperial College London, UK
email: {daniel.jarrett05, e.habets, mark.r.thomas02, p.naylor} @imperial.ac.uk

ABSTRACT

A method is proposed for simulating the sound pressure signals on a spherical microphone array in a reverberant enclosure. The method employs spherical harmonic decomposition and takes into account scattering from a solid sphere. An analysis shows that the error in the decomposition can be made arbitrarily small given a sufficient number of spherical harmonics.

Index Terms— Image method, reverberation, room acoustics, spherical microphone arrays

1. INTRODUCTION

Current and previous studies of spherical microphone arrays either do not take into account the scattering caused by a closed/rigid sphere or else assume free-space conditions. However, at high frequencies, the scattering effect cannot be ignored and in small rooms such as offices and meeting rooms where microphone arrays are expected to become commonplace, there is a non-negligible amount of reverberation. Studies involving real measurements inevitably include both these effects. However, the amount of reverberation is difficult to control and extensive tests with real measurements are expensive (due to the high cost of the arrays) and time-consuming. There is therefore a strong need for a method to simulate room impulse responses between an acoustic source and spherical microphone array in the presence of reverberation.

We propose a method based on spherical harmonic decomposition (SHD). The room reverberation is modelled in a computationally efficient way using a variant of Allen & Berkley's image method [1]. The image method is a widely used method for simulating room impulse responses in small rooms which can in principle be used to simulate responses for microphones on an open sphere. However, practical spherical arrays usually involve microphones placed on a rigid sphere and in this case the image method cannot model the scattering effect of the sphere.

The image method makes a number of assumptions which we also adopt here. The rooms are assumed to be perfectly rectangular which allows for an efficient implementation. The walls of these rooms are assumed to be perfectly rigid, i.e. they have infinite acoustic impedance. Furthermore, when applying the SHD, we assume that the closed sphere is also perfectly rigid. The new method models the direct sound path, reflections due to room reverberation, scattering of the direct path and scattering of the reverberant reflections; we neglect sound scattered and then reflected by the walls of the room, as well as multiple interactions between the sphere and the walls, both of which are considered to be small for a small scatterer [2].

This paper is organized as follows: in Section 2 we introduce the SHD, in Section 3 we explain the impact of sign conventions on the SHD, in Section 4 we present a modified image method which has been adapted to this decomposition, and in Section 5 we analyze the error involved in the SHD. An implementation of the method is available online [3].

2. SPHERICAL HARMONIC DECOMPOSITION

Consider a sound pressure field $p(k, \mathbf{r})$ at a point $\mathbf{r}$ on a sphere, where k is the wavenumber and $\mathbf{r} \triangleq (r, \Omega)$ in spherical polar coordinates (Ω denotes an elevation-azimuth pair). The sound field due to a unit amplitude point source at $\mathbf{r}_{\mathrm{s}} \triangleq (r_{\mathrm{s}}, \Omega_{\mathrm{s}})$ is given by [4, p. 277]

$$
\begin{aligned}
p(k, \mathbf{r}, \mathbf{r}_{\mathrm{s}}) &= 4\pi i k \sum_{l=0}^{\infty} \sum_{m=-l}^{l} b_l(kr) h_l^{(1)}(kr_{\mathrm{s}}) Y_{lm}^*(\Omega_{\mathrm{s}}) Y_{lm}(\Omega) \\
&= 4\pi i k \sum_{l=0}^{\infty} b_l(kr) h_l^{(1)}(kr_{\mathrm{s}}) \sum_{m=-l}^{l} Y_{lm}^*(\Omega_{\mathrm{s}}) Y_{lm}(\Omega),
\end{aligned} \tag{1}
$$

where $(\cdot)^*$ denotes the complex conjugate, $i = \sqrt{-1}$, Y_{lm} is the spherical harmonic of order l and degree m, b_l is the far-field mode strength, and $h_l^{(1)}$ is the spherical Hankel function of the first kind and of order l. Note that as we have chosen a unit amplitude point source, $p(k, \mathbf{r}, \mathbf{r}_{\mathrm{s}})$ is effectively a room transfer function (RTF).

According to the spherical harmonic addition theorem,

$$
\frac{2l+1}{4\pi} P_l(\cos\Theta) = \sum_{m=-l}^{l} Y_{lm}^*(\Omega_{\mathrm{s}}) Y_{lm}(\Omega), \tag{2}
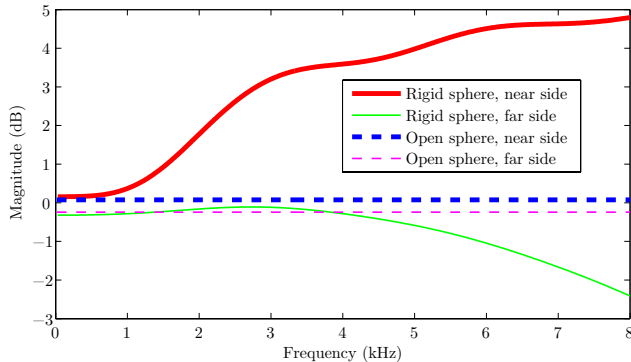$$

Fig. 1. A comparison of RTF magnitudes in the open and rigid sphere cases, for two microphones positioned on a sphere of radius 2 cm: one which is close to the source (near side) and one which is far from the source (far side).

where P_l is the Legendre polynomial and Θ is the angle between Ω_s and Ω. As in [5], (1) therefore reduces to:

$$p(k, \mathbf{r}, \mathbf{r}_s) = ik \sum_{l=0}^{\infty} b_l(kr) h_l^{(1)}(kr_s)(2l + 1) P_l(\cos \Theta). \quad (3)$$

The far-field mode strength for an open sphere is given by $b_l(kr) = j_l(kr)$, where j_l is the spherical Bessel function of order l. For the sound field on the surface of a rigid sphere (such as the *em32 Eigenmike* by mh acoustics), the far-field mode strength includes a scattering term which is derived by assuming that the radial velocity is zero on the surface of the sphere [4, p. 228]:

$$b_l(kr) = j_l(kr) - \frac{j_l'(kr)}{h_l^{(1)'}(kr)} h_l^{(1)}(kr), \quad (4)$$

where $(\cdot)'$ denotes the first derivative. This model of scattering is used in all theoretical analyses for spherical microphone arrays (e.g. [6, 7]), as well as for the design of beamformers which have been experimentally tested [8].

To demonstrate the scattering effect, the magnitude of the RTFs obtained for two microphones positioned on the near side and far side of an array, with open and rigid spheres of radius 2 cm, is shown in Fig. 1, for a source-array distance of 1 m. These RTFs were computed using an implementation [3] of the method in Section 4. While in the case of the open sphere the magnitude response is flat, we see that in the rigid sphere case the magnitude response is frequency-dependent: for the near side microphone there is a constructive effect which increases with frequency, whereas for the far side microphone there is a small constructive effect at low frequencies and an increasingly destructive effect thereafter.

In the case of an open sphere (no scattering), the SHD in (3) is equal to the free space Green's function [4, p. 259]:

$$p(k, \mathbf{r}, \mathbf{r}_s) = \frac{e^{+ik||\mathbf{r} - \mathbf{r}_s||}}{||\mathbf{r} - \mathbf{r}_s||}. \quad (5)$$

3. SIGN CONVENTIONS IN THE WAVE EQUATION

It should be noted that thus far we have considered only RTFs; the sound pressure is in the frequency domain. If we wish to obtain a room impulse response (RIR), we must be careful to use the correct sign convention for the Fourier transform pair. The above expressions all assume the convention for describing waves used in physics where the Fourier transform is defined as

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{+i\omega t} dt, \quad (6)$$

in order to eliminate the $e^{-i\omega t}$ term in the time-harmonic solution to the wave equation. The exponential term in this transform definition has opposite sign to the one habitually used in signal processing, however in the field of acoustics and spherical harmonics it is the more common choice, as in [1, 4]. In this convention, outgoing and incoming waves are respectively described by Hankel functions of the first and second kind.

Alternatively we can use the electromagnetics/engineering convention where the exponential in the Fourier transform definition has opposite sign. The exponential in the Green's function then also has opposite sign, and we must take the complex conjugate of the right-hand side of (3), as in [6].

Using the inverse Fourier transform associated with the forward transform defined in (6), in the open sphere case the RIR is given by the inverse transform of (5):

$$p(t, \mathbf{r}, \mathbf{r}_s) = \frac{\delta(t - \frac{||\mathbf{r} - \mathbf{r}_s||}{c})}{||\mathbf{r} - \mathbf{r}_s||}, \quad (7)$$

where δ is the Dirac delta function and c is the speed of sound.

4. SIMULATION OF IMPULSE RESPONSES

4.1. Method

The image method [1] provides a way of simulating the impulse response between two points in a rectangular room, taking into account reverberation. This method could be used for simulating spherical microphone arrays by considering each of the array's microphones as separate receivers but this would not account for scattering off the rigid sphere, which is the main aim of our work. Our approach models the reflections from the walls of the room similarly to the image method, but uses the image positions in a SHD of the sound field instead of the exponential in (5).

We 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 $v = L_v$.
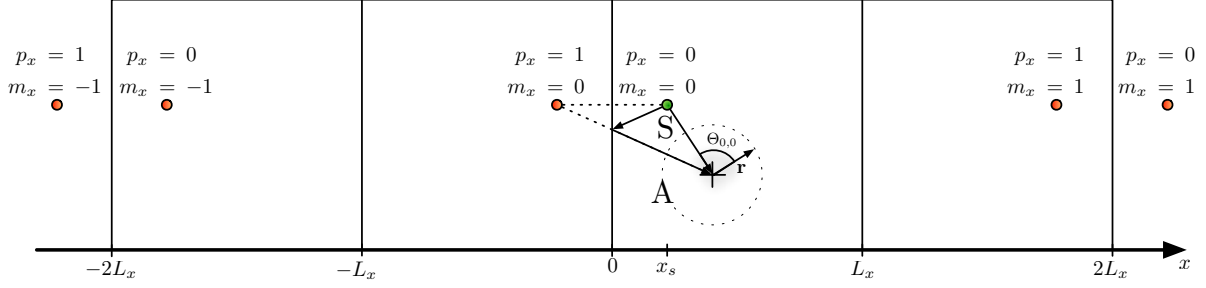
Fig. 2. A slice through the image space showing the positions of the images, with a source S and array A. The full image space has three dimensions (x, y and z). An example of a reflected path is shown for the image with $p_x = 1$ and $m_x = 0$.

If the sound source is located at $\mathbf{r}_s = (x_s, y_s, z_s)$ and the centre 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 $z = 0$ can be expressed as a vector $\mathbf{R}_p$:

$$\mathbf{R}_p = [x_s - 2p_x x_s - x_a, y_s - 2p_y y_s - y_a, z_s - 2p_z z_s - z_a], \quad (8)$$

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 which form a set $\mathcal{P}$. To consider all reflections we also define a vector $\mathbf{R}_m$:

$$\mathbf{R}_m = [2m_x L_x, 2m_y L_y, 2m_z L_z], \quad (9)$$

where each of the elements in $\mathbf{m} = (m_x, m_y, m_z)$ can take values between $-N$ and N, thus resulting in a set $\mathcal{M}$ of $(2N + 1)^3$ combinations. The image positions in the x dimension are illustrated in Fig. 2.

We define $\mathbf{R}_{p,m} \triangleq \mathbf{R}_p + \mathbf{R}_m$. The distance between an image and the centre of the sphere is then given by $||\mathbf{R}_{p,m}||$ and the angle of the image taken with respect to the centre of the sphere is given by $\angle\mathbf{R}_{p,m}$.

In the conventional image method, the direction of the vector $\mathbf{R}_{p,m}$ 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 conventional image method as only the norm of this vector is used. However, for the SHD we also need the angle of the images. Our definition of $\mathbf{R}_p$ in (8) is therefore intentionally different from the corresponding definition in the original image method ([1], equation 6). We define $\mathbf{R}_p$ such that the vector $\mathbf{R}_{p,m}$ always points from the receiver to the image and not in the opposite direction.

If we denote the sound field (including reflections) due to a unit amplitude point source as p_r then, from (3):

$$p_r(k, \mathbf{r}, \mathbf{r}_s) = ik$$
$$\cdot \sum_{\mathbf{p}\in\mathcal{P}} \sum_{\mathbf{m}\in\mathcal{M}} \beta_{x_1}^{|m_x - p_x|} \beta_{x_2}^{|m_x|} \beta_{y_1}^{|m_y - p_y|} \beta_{y_2}^{|m_y|} \beta_{z_1}^{|m_z - p_z|} \beta_{z_2}^{|m_z|}$$
$$\cdot \sum_{l=0}^{\infty} b_l(kr) h_l^{(1)}(k\,||\mathbf{R}_{p,m}||)(2l+1) P_l(\cos\Theta_{p,m}), \quad (10)$$

where $\Theta_{p,m}$ is the angle between $\mathbf{R}_{p,m}$ and $\mathbf{r}$. As we are not interested in $\angle\Theta_{p,m}$, but only $\cos\Theta_{p,m}$, we simply take the dot product of the two normalized vectors:

$$\cos\Theta_{p,m} = \hat{\mathbf{R}}_{p,m} \cdot \hat{\mathbf{r}}, \quad (11)$$

where $\hat{\mathbf{R}}_{p,m} = \frac{\mathbf{R}_{p,m}}{||\mathbf{R}_{p,m}||}$ and $\hat{\mathbf{r}} = \frac{\mathbf{r}}{||\mathbf{r}||}$.

4.2. Implementation

Implementing the proposed method involves evaluating (10) for K discrete values of wavenumber k. We first compute the vector $\mathbf{R}_{p,m}$ and its norm $||\mathbf{R}_{p,m}||$ for all possible $(\mathbf{p}, \mathbf{m})$ values, where the limit N for values (m_x, m_y, m_z) is related to the room dimensions and desired RIR length. We discard images for which $(||\mathbf{R}_{p,m}|| + r)/c$ is larger than the desired RIR length so that when we take the inverse discrete Fourier transform (DFT) of the RTF, the wrap-around effect of the DFT is reduced.

For each microphone position $\mathbf{r}$ we then sum up the SHDs in (3) over all the computed image positions. This requires us to determine the angle between $\mathbf{R}_{p,m}$ and $\mathbf{r}$ as in (11). We can then compute the RTF, and take the inverse DFT to obtain the RIR. Further details are available in our MATLAB implementation [3].

5. FINITE ORDER ERROR

In a computer-based simulation, it is not possible to evaluate the expression for p in (10); the sum over an infinite number of orders l must be replaced with a truncated sum $\hat{p}$ up to a finite order L, i.e., a sum of $L + 1$ terms. We seek to estimate the error involved in this truncation. As (10) is a sum of SHDs with different source positions, the error in the reverberant case is the sum of the errors in the anechoic case. Therefore we restrict our error analysis to the error in (3).

For an open sphere, the error can be straightforwardly determined as the SHD is a decomposition of the closed-form expression in (5). For a rigid sphere, however, no closed-form expression exists since the scattering term can only be
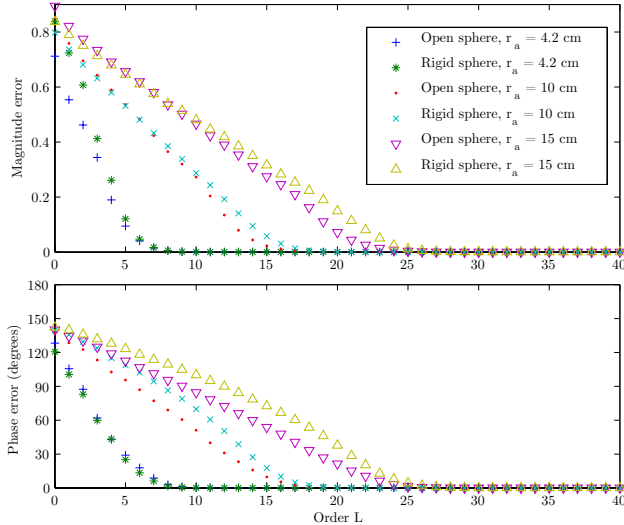
Fig. 3. Finite order errors (magnitude and phase) in the SHD, for rigid and open spheres.

expressed in the spherical harmonic domain. We will therefore estimate the error by comparing the truncated SHD $\hat{p}$ to a high-order (high L) SHD (our reference). 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 SHD as a reference as opposed to the untruncated SHD is small. There are unfortunately numerical difficulties involved in multiplying high order spherical Bessel and Hankel functions, therefore we cannot choose extremely large values of L; L will be chosen to be as large as possible without numerical overflow. For typical sphere radii and source-array distances, this allows us to reach L values of up to about 100 using MATLAB.

We evaluate the SHD at 1024 discrete values of k (denoted by $\tilde{k}$) corresponding to frequencies up to 8 kHz, and then calculate the normalized RMSE (NRMSE) over K values of $\tilde{k}$ in the range 100 Hz - 8 kHz[1] and form a set $\mathcal{K}$, i.e.,

$$\text{NRMSE}_g(\mathbf{r}, \mathbf{r}_\text{s}) =$$

$$\frac{1}{K} \sqrt{\sum_{\tilde{k} \in \mathcal{K}} \frac{\left(g\left[p(\tilde{k}, \mathbf{r}, \mathbf{r}_\text{s}) \right] - g\left[\hat{p}(\tilde{k}, \mathbf{r}, \mathbf{r}_\text{s}) \right] \right)^2}{g\left[p(\tilde{k}, \mathbf{r}, \mathbf{r}_\text{s}) \right]^2}}, \quad (12)$$

where $g(x) = \text{abs}(x)$ for magnitude error and $g(x) = \text{angle}(x)$ for phase error. We average this NRMSE over 32 quasi-equidistant microphones and 50 random source positions at a fixed distance from the centre of the array.

The resulting average errors are given in Fig. 3, for both the open and rigid sphere cases. Three different sphere

---

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

radii are used: $r = 4.2$ cm (the radius of the *Eigenmike*), $r = 10$ 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 yield only a very small reduction in error. A rule of thumb for choosing L is $L > \lceil 1.1\, k_\text{max} r \rceil$ where k_max is the highest wavenumber of interest.

6. CONCLUSIONS

A method has been presented for simulating RIRs for a rigid spherical microphone array in a reverberant environment, enabling algorithms to be rigorously tested by taking into account the combined effects of room reverberation and scattering from the sphere. There is a small error involved in the spherical harmonic decomposition, however this can be controlled at the expense of increased computational complexity.

7. REFERENCES

[1] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.

[2] T. Betlehem and M. A. Poletti, "Sound field reproduction around a scatterer in reverberation," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2009, pp. 89–92.

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

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

[5] E. Fisher and B. Rafaely, "The nearfield spherical microphone array," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2008, pp. 5272–5275.

[6] J. Meyer and G. W. Elko, "Position independent close-talking microphone," *Signal Processing*, vol. 86, no. 6, pp. 1254–1259, Jun. 2006.

[7] J. J. Bowman, T. B. A. Senior, and P. L. E. Uslenghi, Eds., *Electromagnetic and Acoustic Scattering by Simple Shapes*. North-Holland, 1969.

[8] S. Yan, H. Sun, U. P. Svensson, X. Ma, and J. M. Hovem, "Optimal modal beamforming for spherical microphone arrays," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. PP, no. 99, 2010.