

# EFFICIENT EVALUATION OF REVERBERANT SOUND FIELDS

Ramani Duraiswami, Nail A. Gumerov, Dmitry N. Zotkin, Larry S. Davis

Perceptual Interfaces and Reality Laboratory  
Institute for Advanced Computer Studies  
University of Maryland, College Park, MD 20742  
{ramani, gumerov, dz, lsd}@umiacs.umd.edu

## ABSTRACT

An image method due to Allen and Berkley (1979) is often used to simulate the effect of reverberation in rooms. This method is relatively expensive computationally. We present a fast method for conducting such simulations using multipole expansions. For  $M$  real and image sources and  $N$  evaluation points, while the image method requires  $O(MN)$  operations, our method achieves the calculations in  $O(M + N)$  operations, resulting in a substantial speedup. Applications of our technique are also expected in simulation of virtual audio.

## 1. INTRODUCTION

The reverberant characteristics of a room or space plays a large role in determining its acoustic ambience. “Auralization” is a field that aims at determining this characteristic computationally and uses it to render proper sound. On a more practical note, the highly reverberant characteristics of rooms and auditoriums create significant problems for algorithms in speech recognition and acoustical source localization. Thus, a major component of algorithm evaluation in these fields is to test performance under reverberant conditions. Sometimes this evaluation is performed under experimental room conditions. However, since the reverberant response of a room is a function of its geometry, locations of source(s) and receiver(s), the impedance characteristics of walls etc., such experimental evaluations are tedious and potentially expensive. A more economical and usually employed alternative is computer simulation of the reverberant characteristics.

The basis for many computational simulations in both applications is the method is due to Allen and Berkley (1979) [1]. That paper performs an evaluation of reverberative effects of the room on the sound produced by a source using an image method. The effects of the walls are replaced by a system of image sources, comprising both direct and secondary sources, with some heuristic adjustments of the image source strengths, to approximately account for unmodeled effects, such as wall impedances. Using this image system the appropriate impulse response that must be convolved with the source signal can be computed. Reference [1] has been cited in over 80 times in journal papers, and has been cited in an order of magnitude larger number of conference citations (including at least 5 at the recent ICASSP 2001). Citations include papers in microphone array studies [3][4][5], speech [7][6][8], and virtual auditory environments [9][10][11].

The computation of the point-to-point room impulse response using [1] is relatively time consuming for highly reverberant rooms

We would like to gratefully acknowledge the support of NSF award 0086075.

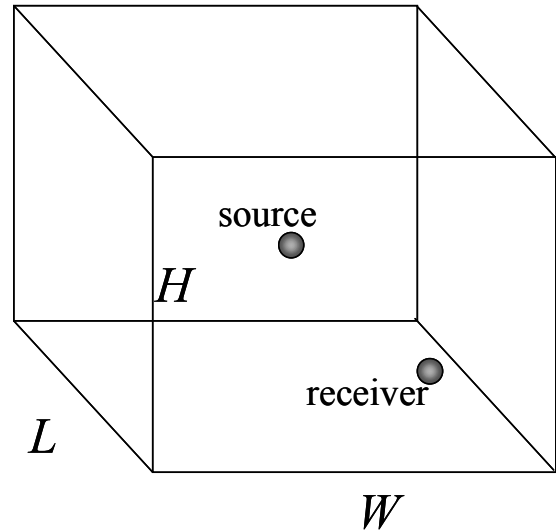


Figure 1: Problem formulation.

and large numbers of evaluation points. The goal of our paper is to introduce a fast algorithm based on multipole expansions that can achieve an order of magnitude improvement in the computational complexity. While the evaluation of the room response function for  $M$  total sources (actual and image) and  $N$  evaluation points using the algorithm of [1] (referred to as the AB algorithm from this point on) requires  $O(MN)$  operations, our proposed method achieves this evaluation in  $O(M + N)$  operations.

The rest of our paper is organized as follows. In Section 2 we describe in brief the problem formulation and the AB algorithm. In Section 3 we describe the multipole algorithm proposed here. Section 4 presents some results of a comparison between our implementation of the AB algorithm and that proposed in this paper. In Section 5 we conclude the paper with possible future applications of the promising method proposed in this paper.

## 2. FORMULATION

Consider a room of length  $L$ , width  $W$ , and height  $H$ . A Cartesian reference frame is connected with the center of this room and axes are made parallel to the walls, so that the 6 walls have coordinates  $x = \pm L/2$ ;  $y = \pm W/2$ ,  $z = \pm H/2$ . Inside the room we have

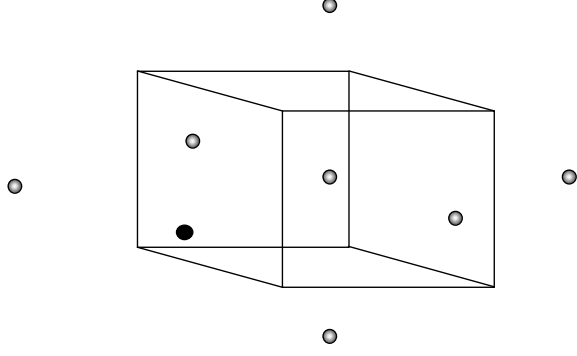


Figure 2: The first six images of the source are shown. The dark circle is the receiver location.

a monopole sound source of intensity  $Q_0$  placed at  $x = x_0, y = y_0, z = z_0$  which in the absence of the walls generates the acoustic field

$$P_0 = \frac{Q_0}{4\pi |\mathbf{r} - \mathbf{r}_0|} e^{ik|\mathbf{r} - \mathbf{r}_0|}. \quad (1)$$

This field satisfies the Helmholtz equation with Sommerfeld radiation conditions,

$$\begin{aligned} \nabla^2 P_0 + k^2 P_0 &= -\delta(|\mathbf{r} - \mathbf{r}_0|), \quad (2) \\ \lim_{r \rightarrow \infty} r \left( \frac{\partial P_0}{\partial r} - ikP_0 \right) &= 0. \end{aligned}$$

Equation (2) is obtained via the Fourier transform of the wave equation. Here  $k = \omega/c$  is the wavenumber,  $\omega$  frequency, and  $c$  the sound speed.

The presence of walls requires the pressure field to satisfy boundary conditions at them. Usually a no flux boundary condition is imposed corresponding to sound hard boundaries.

$$\frac{\partial P}{\partial n} = 0. \quad (3)$$

Other boundary conditions are possible, but are not considered in the formulation of [1].

### 2.1. The AB Algorithm

With a monopole source in a rectangular room, the boundary condition (3) can be modelled by placing image sources which mirror the actual source and all other images with respect to the walls. The location of these image sources is denoted as  $\mathbf{r}_q, q = 1, 2, \dots$ . There will be an infinite number of sources to model reflection of signals of arbitrary frequency and duration.

In practice the walls are not ideal, and absorb part of the sound and have impedance. However in the AB formulation this is dealt with by modifying the strengths of image sources by a factor  $\beta_i$  corresponding to reflections in each direction.

Further, the duration of acoustic signals is limited. We denote  $t_{\max}$  the maximum time considered. Since acoustic waves propagate with finite speed only image sources with  $|\mathbf{r}_q - \mathbf{r}| \leq ct_{\max}$  influence the sound field inside the room, where  $\mathbf{r}$  is the radius vector of an arbitrary point inside the room.

Let  $D_{\min} = \frac{1}{2}\sqrt{L^2 + W^2 + H^2}$  be the distance from the center of the room to any corner (vertex of the rectangular box).

Let  $S_{\min}$  be the sphere with radius  $D_{\min}$  centered at the origin of the reference frame. This sphere includes the entire room. Consider also a sphere  $S_{\max}$  of radius  $R_{\max} = ct_{\max} + D_{\min}$  concentric to  $S_{\min}$ . Assume that we have  $N_{\max}$  image sources located inside  $S_{\max}$ . The image sources located outside  $S_{\max}$  do not influence the sound field in the room for time interval  $[0, t_{\max}]$ . Therefore the field inside the room within specified time limits can be represented as

$$P(\mathbf{r}; k) = \sum_{|\mathbf{r}_q| < R_{\max}} P_q(\mathbf{r}; k), \quad q = 0, \dots, N_{\max} \quad (4)$$

where

$$P_q(\mathbf{r}) = \frac{Q_q}{4\pi |\mathbf{r} - \mathbf{r}_q|} e^{ik|\mathbf{r} - \mathbf{r}_q|}, \quad (5)$$

$Q_q$  is the strength of the  $q$ th source. For sound-hard walls the intensity of the image sources is the same as the intensity of the real source,  $Q_q = Q_0$ . In the case of absorbing walls  $Q_q$  is set to be  $Q_q = \beta_q Q_0$ , where the coefficients  $\beta_q$  model sound absorption. As discussed in [1] this corresponds to the assumption of a non-physical directional absorption model. Despite this criticism it has been used by several authors to model room acoustics.

Formula (4) provides the solution of the problem. The impulse response of the room can be calculated by taking the inverse Fourier transform.

$$p(\mathbf{r}; t) = \sum_{|\mathbf{r}_q| < R_{\max}} \frac{Q_q \delta\left(t - \frac{|\mathbf{r} - \mathbf{r}_q|}{c}\right)}{4\pi |\mathbf{r} - \mathbf{r}_q|}. \quad (6)$$

Thus, given a broadband source  $S(t)$  we can convolve it with the room impulse response corresponding to the source and receiver locations.

### 3. MULTIPOLE METHOD

The practical shortcoming of this solution is that the number of image sources within  $S_{\max}$  can be large (several thousand) and evaluation of the sum over  $N$  sources in  $S_{\max}$  for any given location in the room can be an expensive computational procedure. Further, we may need the solution to be evaluated at many points, say  $M$ , in the room (for example in an auralization calculation). In general this requires computation of  $O(NM)$  distances and subsequent multiplications and additions.

We propose and implement a new method for evaluation of this sum, using a multipole reexpansion technique. This provides a fast algorithm requiring  $O(N + M)$  operations for evaluation of the sound field.

We propose to evaluate  $P(\mathbf{r})$  in the following way. We break the sum (4) into two terms:

$$P(\mathbf{r}) = \psi_s(\mathbf{r}) + \psi_r(\mathbf{r}), \quad (7)$$

where  $\psi_s$  and  $\psi_r$  are respectively the singular and regular parts of the field inside the sphere  $S_D$  of radius  $D \geq D_{\min}$  concentric to  $S_{\min}$  and are given by

$$\psi_s(\mathbf{r}) = \sum_{r_q < D} \psi_q(\mathbf{r}), \quad \psi_r(\mathbf{r}) = \sum_{D < r_q < R_{\max}} \psi_q(\mathbf{r}), \quad (8)$$

The singular part is generated by the real source, plus image sources  $\psi_q(\mathbf{r})$  located inside the sphere  $S_D$ . The regular part is generated

by the image sources located in the domain bounded by spheres  $S_D$  and  $S_{\max}$ . Below we consider evaluation of the regular part.

A monopole source centered at the origin, can be expanded into a series of spherical harmonics centered about a point  $q$  using the following identity (see e.g., the text of Morse & Feshbach [13])

$$\psi_q(\mathbf{r}) = ik \sum_{n=0}^{\infty} \sum_{m=-n}^{m=n} S_n^{-m}(\mathbf{r}_q) R_n^m(\mathbf{r}), \quad r \leq r_q, \quad (9)$$

where  $r_q$  denotes the distance from point  $q$ , and

$$\begin{aligned} R_n^m(\mathbf{r}) &= j_n(kr) Y_n^m(\theta, \varphi), \\ S_n^{-m}(\mathbf{r}_q) &= h_n(kr_q) Y_n^{-m}(\theta_q, \varphi_q), \\ Y_n^m(\theta, \varphi) &= \sqrt{\frac{2n+1}{4\pi} \frac{(n-|m|)!}{(n+|m|)!}} P_n^{|m|}(\cos \theta) e^{im\varphi}. \end{aligned} \quad (10)$$

Here  $(r, \theta, \varphi)$  are spherical coordinates of the radius vector  $\mathbf{r}$ ,  $j_n(kr)$  and  $h_n(kr)$  are the spherical Bessel and Hankel functions, and  $Y_n^m(\theta, \phi)$  are the orthonormal spherical harmonics, and  $P_n^m(\mu)$  are the associated Legendre functions.

Thus, the regular part of the field can be written in the form:

$$\psi_r(\mathbf{r}) = ik \sum_{n=0}^{\infty} \sum_{m=-n}^{m=n} C_n^m R_n^m(\mathbf{r}), \quad (11)$$

$$C_n^m = \sum_{D < r_q < R_{\max}} Q_q S_n^{-m}(\mathbf{r}_q), \quad (12)$$

where the coefficients  $C_n^m$  can be precomputed for given image source locations. These series are very quickly convergent and can be truncated at  $s$  terms, and evaluation of  $\psi_r(\mathbf{r})$  at any given point  $\mathbf{r}$  can be restricted to the first  $s$  terms, thereby requiring a summation of  $m = (s+1)^2$  terms. This must be compared with summation over  $N$  terms due to the sources with  $D < r_q < R_{\max}$  if formula (8) were used directly. Note that  $m$  only depends on the desired accuracy of the summation, and can be orders of magnitude smaller than  $N$ . Thus the evaluation only requires  $O(M+N)$  operations with the constant depending on the desired accuracy. This gain is more significant when the number of evaluation points,  $M$ , where  $\psi_r(\mathbf{r})$  must be evaluated is large.

A sketch demonstrating conceptually the computational complexity comparison of the AB and the Multipole based algorithms is shown in Figure 3. In this figure  $l$  denotes the number of operations for evaluation of a single source for a given field point. Accounting for contribution of a single source to the multipole coefficients requires  $n$  operations which is of the same order as  $m$ . Here  $n$  and  $m$  depend on the desired accuracy.

### 3.1. Error Evaluation

Direct numerical implementation of formula (11) shows a problem related to the fact that at large  $n$  and fixed  $\mathbf{r}_q$  the magnitude of  $S_n^{-m}(\mathbf{r}_q)$  grows as  $n^n$ . This imposes limitations on the order at which we can truncate the series. For large  $n$  and fixed  $\mathbf{r}$  and  $\mathbf{r}_q$  the principal terms of asymptotic expansions of  $S_n^{-m}(\mathbf{r}_q)$  and  $R_n^m(\mathbf{r})$  (see Equation 10) can be found using asymptotic expressions for Bessel functions from Abramowitz & Stegun [12]

$$\begin{aligned} S_n^{-m}(\mathbf{r}_q) &\sim -\frac{i}{kr_q} \sqrt{\frac{2}{e}} \left( \frac{2n+1}{ekr_q} \right)^n Y_n^{-m}(\theta_q, \varphi_q), \\ R_n^m(\mathbf{r}) &\sim \frac{1}{2n+1} \sqrt{\frac{e}{2}} \left( \frac{ekr}{2n+1} \right)^n Y_n^m(\theta, \varphi). \end{aligned} \quad (13)$$

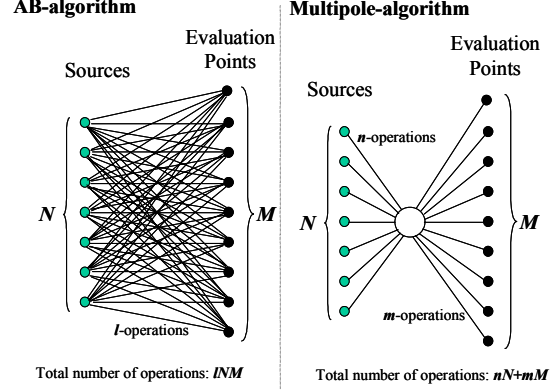


Figure 3: Comparison of operation counts for the direct evaluation using the algorithm of [1] (left) and the proposed algorithm (right).

These expressions allow one to determine the order of truncation. Let us define

$$s_k(r) = \frac{ekr-1}{2}. \quad (14)$$

This has the following meaning: for given  $r_q$  and  $n > s_k(r_q)$  functions  $S_n^{-m}(\mathbf{r}_q)$  grow explosively. At the same time the product  $S_n^{-m}(\mathbf{r}_q) R_n^m(\mathbf{r})$  is finite and converging fast to zero with increasing  $n$  for any  $r/r_q < 1$ . It can be evaluated for large  $n$  and fixed  $\mathbf{r}$  and  $\mathbf{r}_q$  using (13)

$$S_n^{-m}(\mathbf{r}_q) R_n^m(\mathbf{r}) = \left( \frac{r}{r_q} \right)^n \frac{Y_n^{-m}(\theta_q, \varphi_q) Y_n^m(\theta, \varphi)}{ikr_q(2n+1)}. \quad (15)$$

This evaluation shows that the truncation error decreases with  $n$  geometrically. It also depends on the ratio  $r/r_q$ , which decays for sources located far from the evaluation points. The maximum value of  $r/r_q$  is realized for sources at the distance about  $r_q \sim D$ . Therefore, selection of truncation number as

$$s = s_k(D) \quad (16)$$

provides both stable computation of the series for the entire domain and provides the following evaluation of the truncation error for any point inside the room

$$\epsilon_s = O\left( \frac{1}{(2s+1)^2} \left( \frac{D_{\min}}{D} \right)^s \right). \quad (17)$$

## 4. RESULTS AND DISCUSSION

To check the accuracy and performance of the proposed algorithm we implemented it using Matlab and compared results with those obtained using the AB algorithm. Figure 4 shows an example comparison between results of computations. In this example the room size was  $2\text{m} \times 2\text{m} \times 2\text{m}$  and evaluation points were distributed over the entire room. The radius  $R_{\max}$  was about 12 m, which corresponds to  $t_{\max} = 30$  ms, and includes the monopole source and its 910 images. The wavenumber  $k$  was  $3 \text{ m}^{-1}$  (corresponds to 164 Hz).

In computations we selected  $D = D_{\min} = 1.73$  m and varied the truncation number  $s$ . Results shown in Figure 4 were obtained

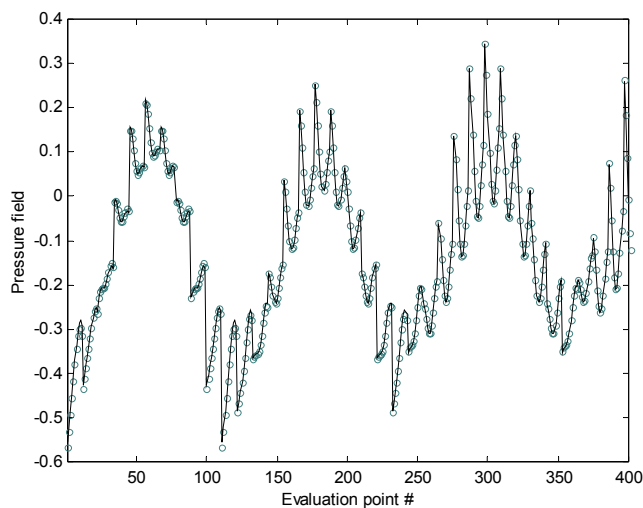


Figure 4: Comparison of the pressure field in a  $2\text{m} \times 2\text{m} \times 2\text{m}$  reverberant room using the algorithm of Ref. [1] (solid curve) and the proposed Multipole algorithm (circles) for 400 evaluation points ( $k = 3 \text{ m}^{-1}$ ,  $t_{\text{max}} = 30 \text{ ms}$ ). No difference can be observed.

for  $s = s_k(D) = 7$ . Even for low values of  $s$  the accuracy of computations is remarkable. Use of large values of  $s$  does not make sense, since it does not substantially improve accuracy, requires larger computation times and may cause computational problems due to the above mentioned exponential growth of coefficients of multipole expansion  $C_n^m$  (11) at large  $n > s_k(D)$ .

The computational time for the case discussed for  $N = 1331$  evaluation points was about 2.7 times smaller when using the proposed algorithm compared to the AB-algorithm. As it was emphasized above the gain in computational time increases with the increase of  $M$  and  $N$ . For example for  $R_{\text{max}} = 19 \text{ m}$ , which corresponds to  $t_{\text{max}} = 50 \text{ ms}$ , and includes the source and its 3534 images, the Multipole algorithm performed the same evaluations 6.5 times faster than the AB algorithm. It should be noted that our computations of special functions are in no way optimized and use time consuming recursions. This performance is consistent with the evaluation of the number of operations for these algorithms. Development of optimized versions of our routines should yield further gains.

## 5. CONCLUSIONS

We proposed and tested an algorithm for fast computation of room acoustics based on multipole expansions. Comparisons with the AB-algorithm [1] showed that computations with similar accuracy can be performed at substantial speed-up.

It is our hope that use of such multipole methods (e.g., in conjunction with boundary element methods) will yield physically valid "auralization" simulations based on the solution of the wave equation, while having complexity similar to more approximate techniques such as those used in [9]-[11].

## 6. REFERENCES

- [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] J. Borish, Extension of the Image Model to Arbitrary Polyhedra. *J. Acoust. Soc. Am.*, 75, 1827-1836, 1984.
- [3] 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.
- [4] T.D. Abhayapala, R.A. Kennedy, R.C. Williamson, "Noise modeling for nearfield array optimization," *IEEE Signal Proc. Let.*, 6, 210-212, 1999.
- [5] M.S. Brandstein, "Time-delay estimation of reverberated speech exploiting harmonic structure," *J. Acoust. Soc. Am.* 105, 2914-2919, 1999.
- [6] A. Spalanzani, H. Kabre, "Evolution, learning and speech recognition in changing acoustic environments," *Lect. Notes Comput. Sc.*, 1498, 663-670, 1998.
- [7] M. Omologo, P. Svaizer, M. Matassoni, "Environmental conditions and acoustic transduction in hands-free speech recognition," *Speech Commun.*, 25, 75-95, 1998.
- [8] P.W. Shields, D.R. Campbell, "Intelligibility improvements obtained by an enhancement method applied to speech corrupted by noise and reverberation," *Speech Commun.* 25, 165-175, 1998.
- [9] T. Funkhouser, I. Carlbom, G. Elko, G. Pingali, M. Sondhi, and J. West. "A Beam Tracing Approach to Acoustic Modeling for Interactive Virtual Environments." *Proc. SIGGRAPH 98*, Orlando, FL, 21-32, 1998.
- [10] L. Savioja, J. Huopaniemi, T. Lokki T, "Creating interactive virtual acoustic environments," *J Audio Eng. Soc.*, 47, 675-705, 1999.
- [11] N. Tsingos, T. Funkhouser, A. Ngan, I. Carlbom, "Modeling Acoustics in Virtual Environments Using the Uniform Theory of Diffraction," *Proc. SIGGRAPH*, 2001.
- [12] M. Abramowitz & I.A. Stegun. *Handbook of Mathematical Functions*, National Bureau of Standards, Wash., D.C., 1964.
- [13] P.M. Morse & H. Feshbach. *Methods of Theoretical Physics -I*, McGraw-Hill, 1953.