Fast evaluation of the room transfer function using the multipole method

Ramani Duraiswami, Dmitry N. Zotkin*, Nail A. Gumerov
Perceptual Interfaces and Reality Laboratory, UMIACS
University of Maryland, College Park, MD 20742 USA
ramani@umiacs.umd.edu, dz@umiacs.umd.edu, gumerov@umiacs.umd.edu
Phone: (301)405-8753
Fax: (301)405-6707

Abstract

Reverberation in rooms is often simulated with the image method due to Allen and Berkley (1979). This method has asymptotic complexity that is cubic in terms of the simulated reverberation length. When employed in the frequency domain, it is relatively computationally expensive to use for many receivers in the room or in a dynamically changing configuration due to the repeated summation of the fields generated by a large number of image sources. In this paper, a fast method to perform such summations is presented. The method is based on the multipole expansion of the monopole source potential. For offline computation of the room transfer function for N image sources and M receiver points, use of Allen-Berkley algorithm requires $O(NM)$ operations, whereas use of the proposed method requires only $O(N+M)$ operations, resulting in significantly faster computation of reverberant sound fields. The proposed method also has considerable speed advantage in situations where the room transfer function must be rapidly updated online in response to the source/receiver location changes. Simulation results are presented, and algorithm accuracy, speed, and implementation details are discussed. For problems that require frequency-domain computations, the algorithm is found to generate sound fields that are identical to the ones obtained with the frequency-domain version of the Allen-Berkley algorithm at a fraction of computational cost.

Index Terms

EDICS Categories: 2-ROOM, 1-ENHA, 2-SMCA.

Some of the results in this paper were presented in abbreviated form in [1]. Partial support of NSF awards 0086075 and 0219681 is gratefully acknowledged. We also thank three anonymous reviewers for their helpful and careful comments and suggestions.

I. Introduction

In the typical inhabited environment, the sound reaching the receiver is not an exact copy of the sound emitted by the transmitter. Instead, the sound is scattered off various surfaces in the environment as well as off the transmitter and the receiver (if they have finite extent). The collective effect of the sound interaction with room walls is known as reverberation. The formal studies of reverberation date back to 1900 [2]. Since then, numerous studies on how reverberation affects human perception of sound and music have been conducted. Several distinct research directions can be noted, such as optimization of the room response for perceptually pleasant listening experience, for the best performance of a human subject doing a specific auditory task (such as localization of an audio source or speech comprehension), and for the best performance of a certain computer algorithm (such as localization and enhancement of an active sound source in the room). The first of these three directions has evolved into a separate branch of acoustics called architectural acoustics (e.g., [3], [4]). Example studies for the second and the third directions mentioned are [5], [6], and [7], [8], respectively.

The reverberation properties of a room play extremely important role in determining the listening experience. For example, in architectural acoustics there exist certain principles for constructing performance spaces to assure that the room acoustics helps the perception of the performance rather than ruining it. In virtual acoustics and audiovisual environments the simulation of reverberation is necessary to assist with externalizing of sounds to achieve a sense of presence and to convey information about the virtual environment properties (room size, surface material, etc.) In teleconferencing it is necessary to account for the reverberant properties of the room to compensate for the acoustic feedback and to improve the accuracy of the localization of the active speaker (e.g., for pointing a pan-tilt camera at the speaker). Finally, enhanced spatially selective sound capture can be performed with knowledge of the sound source position and room geometry by not only focusing the acoustic capture unit on the main acoustic source but also using its coherent reflections from the room walls (a technique known as matched filter array (MFA) processing).

In many studies, it is necessary to perform validating tests under reverberant conditions. Sometimes these are done in an actual experimental room. However, since the reverberant response of a room is a function of its geometry, of the source and the receiver locations, of impedance characteristics of the room walls, and of other parameters, it might be an expensive project to perform a comprehensive experimental evaluation. A more economical, and usually employed, alternative is computer simulation of the room reverberation characteristics. Methods for modeling of the room reverberation have many uses in various application areas where prediction of a room response to an arbitrary stimulus is necessary.

One of the first artificial reverberators aimed at reproducing perceptual characteristics of reverberation

was developed by Schroeder [9] and later greatly improved by Moorer [10]. These however did not attempt to model exact sound propagation paths but rather had the goal of creating a perception similar to the one obtained in a real room. The first geometric model of room reverberation is due to Allen and Berkley [11]. This paper became the basis for many sequential papers (e.g., [12], [13], [14], [15]) and is heavily cited in journals and conferences, including citations in microphone array studies [16], [17], speech processing [18], [19], and virtual auditory environments [20], [21], [22]. The basic idea of the Allen-Berkley model is that the reverberation can be represented as the effect of an infinite number of image sources that are created by reflecting the true acoustic source in room walls. The walls are assumed to be specular reflectors of the sound, the wall reflection coefficients are real and frequency-independent, and the room has a simple rectangular shape, which greatly simplifies the geometric computation of the image source positions. Under these conditions, a simple time-domain summation of contributions of each image source with appropriate strengths gives the simulated room impulse response (RIR). The Fourier transform of a RIR is commonly called room transfer function (RTF). The model was extended in [12] to the case of arbitrary piecewise-planar rooms and in [14] to the case of directional and/or shadowing source and/or receiver. Additionally, it was noted that it is generally unnecessary to model the reverberation as accurately as possible beyond a certain time (about 80 ms) after the arrival of the direct sound [23], [24], and consequently approximations of a statistical nature aimed at evoking the appropriate perception are usually employed to model later parts of the reverberation (e.g., [9], [25], [26]).

As mentioned before, the Allen-Berkley model operates in the time domain with significant simplifying assumptions. It is known that not for all rooms is the reverberation time independent of frequency [14], and that the reflective properties of materials do depend on frequency [27]. In addition, sound absorption by the air is frequency-dependent [28]. The desire to extend the model to allow for independent computations for different frequency bands is expressed in several papers [13], [14], [29]. In some more recent papers [6], [30], fine-band RTF modeling is performed and simulation results in different bands are assembled together to form the final response. As a particular example, MFA processing [31], [32], [33] relies on fast RIR computation in an attempt to coherently pick up not only the source signal itself but also its significant reflections by filtering the received signal with the time reverse of the RIR for given positions of the source and the receiver, resulting in a significant improvement of the output signal SNR compared to simple beamforming. The image model [11] is used to model the RIR in [33]. Application of MFA to real scenarios (such as teleconferencing) calls for a dynamic RIR update in response to speaker motion, and while the time-domain image model is sufficiently fast to enable real-time implementation, it does not account for frequency-dependent wall reflection coefficients which can significantly distort the RIR and consequently the beamformed signal.

There exist two distinct situations where the room transfer function must be computed. In the first case, commonly called offline case, the RTF is to be computed for many sources in the acoustic scene and many receiver locations. Direct application of the Allen-Berkley algorithm in this case results in the computational complexity that grows as the product of the number of sources and the number of receiver locations at which the RTF is desired. In the second (online) case, it is necessary to dynamically compute the RTF for a fixed location of the acoustic source and varying receiver location with a minimal latency, and that latency is the main performance measure (i.e., time spent in offline computations that are done beforehand to prepare any necessary auxiliary data structures can be ignored). The problem can also be posed for the fixed receiver (e.g., a microphone in the microphone array) and moving source by using the reciprocity principle [34]. The goal of current paper is to introduce a fast, frequency-domain, multiple-receivers RTF computation algorithm based on multipole expansion that accelerates computation in both of these situations, with greater benefits in online case.

The algorithm is based on exploitation of the fact that the Allen-Berkley RTF solution, which is nothing but the sum of the potential of many monopole sources, can be represented via the addition theorem in factorized form. Similar factorizations are also used in the fast multipole method, first introduced in [35] and now widely used to accelerate computations for different problems, including boundary element method [36], [37], multiple scattering from many objects [38], and others. The algorithm can achieve several orders of magnitude improvement in the computational complexity in cases when the frequency-domain RTF is needed. If the total number of (actual and image) sources is N and the number of RTF evaluation points is M, then the image model [11] requires $O(NM)$ time, whereas the proposed method takes only $O(N+M)$ time by initially assembling the vast majority of the image sources into one complex multipole source located in the center of the room (the assembly process does not require knowledge of the position of evaluation points) and then evaluating the multipole source potential at the evaluation points. Furthermore, the complexity of the multipole source is directly proportional to the size of the room in which the multipole source is used to represent the sum of potentials of image sources; therefore, the RTF evaluation in online scenario can be done arbitrarily fast with the help of the room subdivision technique covered in Section 6 of this paper.

The rest of the paper is organized as follows. In Section 2, a brief overview of the problem is given and the Allen-Berkley algorithm [11] is discussed. In Section 3, the proposed multipole expansion algorithm is described. In Section 4, the error analysis is given and the guidelines for the choice of the implementation parameters are presented. In Section 5, the computation complexity of the Allen-Berkley algorithm and the proposed algorithm is analyzed. In Section 6, a modification of the algorithm that is suitable for real-time implementation as well as for raising the high frequency limit is proposed. In Section 7, comparisons

between the results computed using Allen-Berkley method and computed using proposed method are shown. In Section 8, the performance and implementation considerations are discussed. Finally, Section 9 concludes the paper.

II. Formulation and the Allen-Berkley algorithm

Consider a room of dimensions $l \times w \times h$ (length, width, and height, respectively). 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 $x = \pm l/2$, $y = \pm w/2$, $z = \pm h/2$. A monopole source of strength Q_0 is placed inside the room at $x = x_0, y = y_0, z = z_0$. In the absence of walls, the source generates the acoustic field

$$\psi_0(\mathbf{r}; k) = \frac{Q_0}{4\pi \left|\mathbf{r} - \mathbf{r}_0\right|} e^{ik|\mathbf{r}-\mathbf{r}_0|}, \tag{1}$$

where $\mathbf{r}_0 = (x_0, y_0, z_0)$ is a vector of the source position. This field satisfies the Helmholtz equation with Sommerfeld radiation condition,

$$\begin{aligned} \nabla^2\psi_0 + k^2\psi_0 &= -\delta\left(|\mathbf{r} - \mathbf{r}_0|\right), \\ \lim_{r\to\infty} r\left(\frac{\partial\psi_0}{\partial r} - ik\psi_0\right) &= 0. \end{aligned} \tag{2}$$

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

$$\frac{\partial\psi}{\partial n} = 0. \tag{3}$$

(These boundary conditions result in non-decaying sound field, which is treated in [11] by heuristically introduced wall absorption. Other boundary conditions are also possible, but are not considered in [11]).

With a monopole source in a rectangular room, the boundary condition (3) results in a specular reflection of sound waves. A reflection of the sound from the wall can equivalently be represented by placement of a monopole image source behind the wall along the perpendicular from the source to the wall so that the distances from the wall to the actual source and to the image source are equal (Figure 1). The waves these image sources emit are in turn reflected off other walls, creating higher order reflections, which are in turn represented by further image sources. The location of these image sources is denoted as $\mathbf{r}_q$, $q = 1, 2, ...$, and there are an infinite number of them. Simple geometric relationships can be employed to find the image source positions. The computations are especially easy for the rectangular-shaped room considered here, in which case a simple, regular, infinite lattice is formed by the image sources (Figure 2 shows an illustration, drawn in two dimensions for simplification).

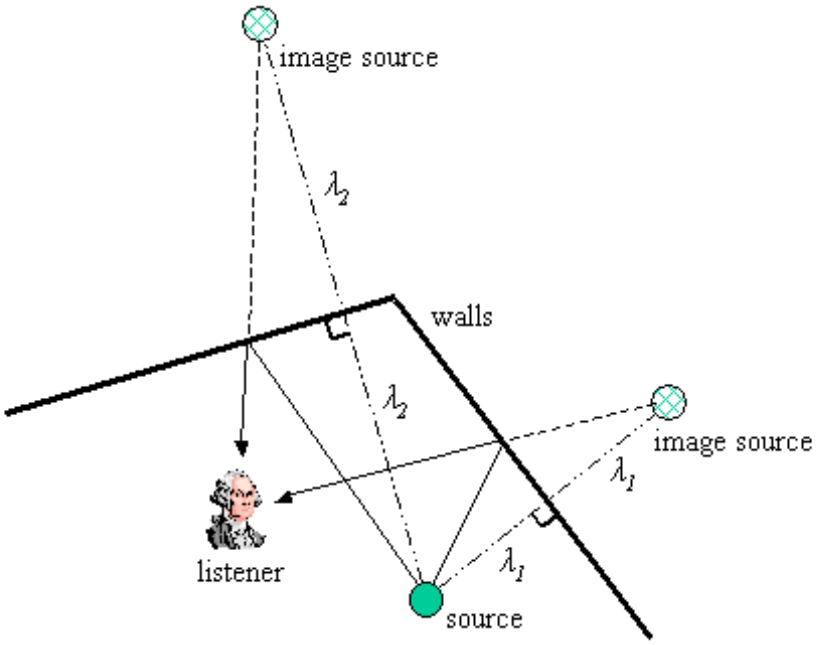

Fig. 1. An illustration to formation of image sources.

In practice the walls are not ideal but rather have impedance and therefore absorb part of the sound. In [11] wall absorption is simulated by introducing six heuristic frequency-independent real-valued wall reflection coefficients $\beta_i, 0 \leq \beta_i < 1$, $i = 1..6$ (one per room wall), and the strength Q_q of the qth image source is obtained by multiplying the strength of the actual source Q_0 (which can be taken to be unit, $Q_0 = 1$) by the reflection coefficients for all walls in which the actual source was reflected to create that image source. As discussed in [11], such a heuristic approach corresponds to the assumption of a nonphysical directional absorption model. Despite this criticism, it has been used by several authors to model room acoustics. (Another, better physically-justified approach is to make the reflection coefficient complex and dependent on the angle of incidence at each wall reflection; in fact, it is necessary for the reflection factor to be angle-dependent to correctly simulate the interaction between the plane wave and the surface [4]. This approach is not followed here, but the same algorithm can be applied in such situation).

It follows that because of the wall absorption the strength of the generated image sources decreases as more and more orders of reflections are considered. Furthermore, in any practical system the reverberation duration must be limited. Let us denote the maximum time considered by $t_{\max}$. Since acoustic waves propagate with finite speed, only image sources that lie within $|\mathbf{r}_q - \mathbf{r}| \leqslant ct_{\max}$ influence the sound field in 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 room center to any room corner (a 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 $R_{\max} = ct_{\max} + D_{\min}$

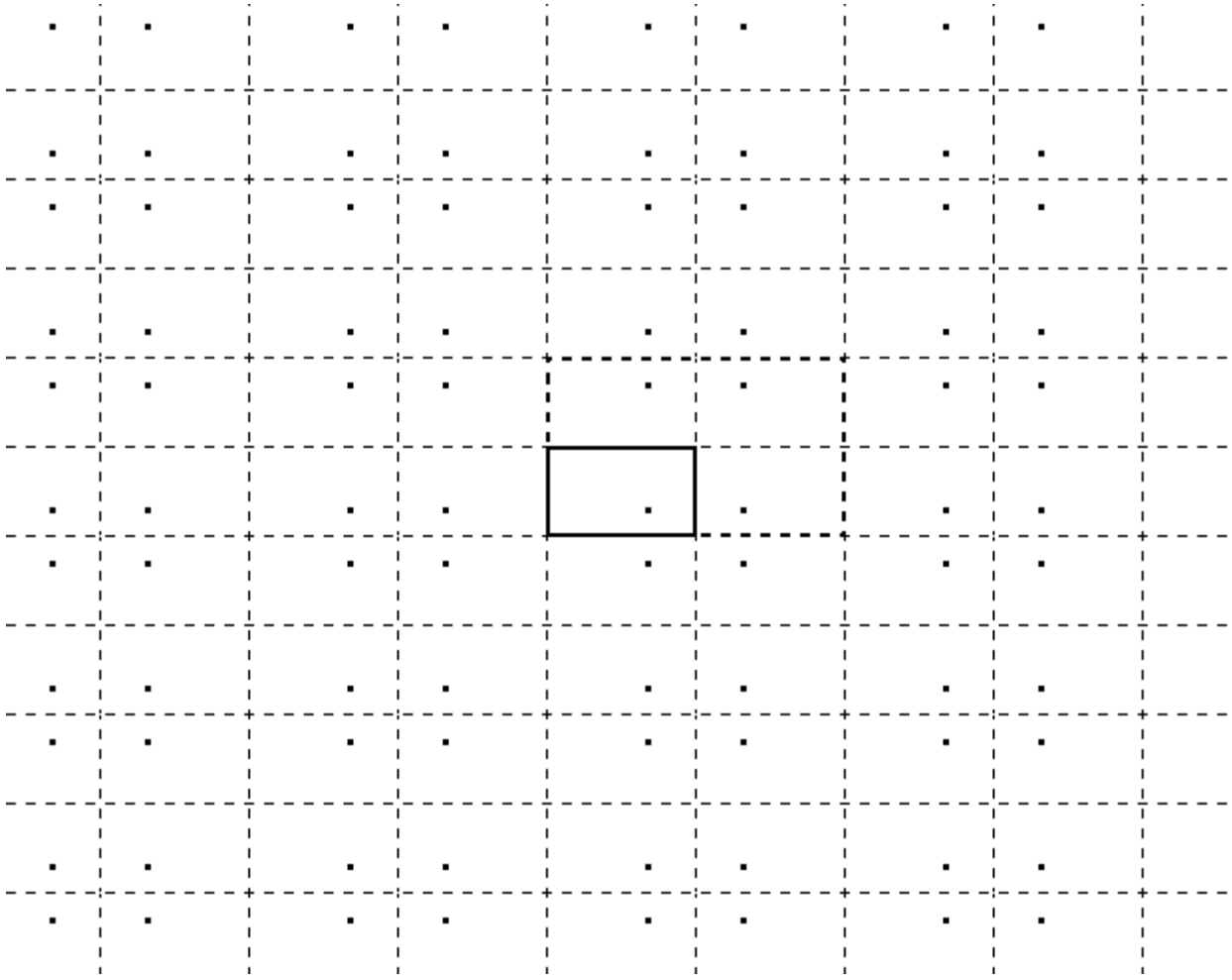

Fig. 2. An illustration to formation of the infinite lattice of image sources in case of the rectangular room. The room is shown as a solid rectangle in the middle, and the black dot inside the room is the actual sound source. All other dots are image sources. The lattice is formed by infinitely repeating the block surrounded by thick dashed line.

concentric to $S_{\min}$. Assume that there are total of $N_{\max}$ 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

$$\psi(\mathbf{r}; k) = \sum_{|\mathbf{r}_q| < R_{\max}} \psi_q(\mathbf{r}; k), \quad q = 0, ..., N_{\max} \tag{4}$$

where

$$\psi_q(\mathbf{r}; k) = \frac{Q_q}{4\pi |\mathbf{r} - \mathbf{r}_q|} e^{ik|\mathbf{r}-\mathbf{r}_q|}, \tag{5}$$

Equation (4) provides the solution of the problem in the form of RTF. The RIR can be calculated by taking the inverse Fourier transform,

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

In fact, this expression accurately describes, up to the scale factor, the process of assembling the time-domain impulse response in [11]. Thus given a broadband source $S(t)$ one can convolve it with the RIR corresponding to the source and the receiver locations.

Since the development of the Allen-Berkley method, much progress was achieved in room reverberation simulation. It is well-known now that more sophisticated methods that are able to handle complex non-rectangular rooms, sound diffraction, diffuse reflections, and grazing incidence effects must be used to simulate sound fields that would agree with the measured RIRs with good accuracy [39], [40], [41]. Also, statistical ray-tracing methods (e.g., [42]) and beam-tracing algorithms (e.g., [20]) are often employed instead of image-based methods. Still, despite its shortcomings, Allen-Berkley algorithm is simple, can be implemented quickly, compares favorably in terms of cost with commercially available room simulation software, and (most important) produces simulated reverberation that is often perfectly acceptable, depending on the purpose of simulation. Because of that, it is still in broad use and is being widely cited by the research community. In this paper, the re-formulation of the algorithm aimed at reducing its computational complexity (especially in case of multiple sources and/or multiple receivers presented in the room) is provided.

III. Multipole method

Despite its computational complexity growing as the product of the number of sources and receivers, in practice, the time-domain solution (6) is fast. However, it assumes no variations in the spectrum of the image sources. To account for possibly non-flat image source frequency content, it is necessary to implement the Allen-Berkley algorithm in the frequency domain by using the equation (4) for many wavenumbers (in fact, half as many as there are points in RIR) to obtain the RTF and then take inverse Fourier transform to get the RIR. In this section, an algorithm to speed up a frequency-domain implementation of the Allen-Berkley method is presented.

The practical shortcoming of the solution (4) is that the number of image sources within $S_{\max}$ can be large (tens of thousands) and evaluation of the sum over N sources in $S_{\max}$ for a given location in the room can be an expensive computational procedure. Furthermore, in some situations the RTF has to be computed at M points in the room. Direct implementation of (4) requires that for each of the M evaluation points N sources be examined and that a source-receiver distance be computed for each pair of the source and the receiver, resulting in a total of $O(NM)$ operations.

A fast method for evaluation of the sum (4) is proposed and implemented here. The method uses a multipole expansion technique and requires only $O(N+M)$ operations for the sound field evaluation.

First, the sum (4) is broken into two terms:

$$\psi\left(\mathbf{r}\right) = \psi_{s}(\mathbf{r}) + \psi_{r}(\mathbf{r}), \tag{7}$$

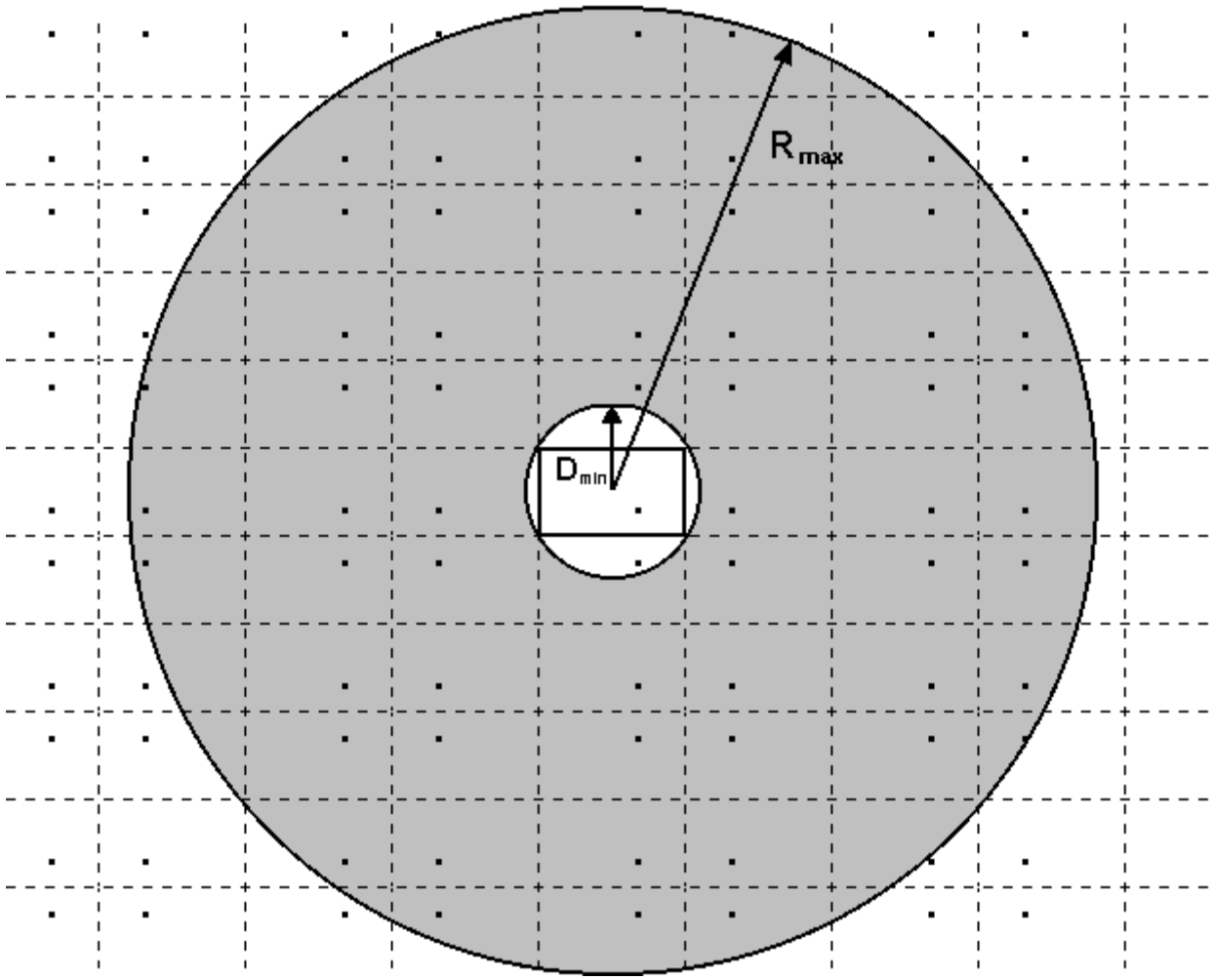

Fig. 3. Part of the infinite lattice of the image sources. The sphere of the radius $D_{\min}$ encloses the room. The sphere of the radius $R_{\max}$ represents the maximum radius of interest, and image sources farther away are ignored. All the image sources that are located in the grey area between two spheres constitute the regular part of the potential. The actual source in the room and one image source that is located outside of the room but inside the smaller sphere constitute the singular part of the potential.

(the dependence on k is suppressed from now on for clarity), where ψ_s and ψ_r are the singular and the regular parts, respectively, of the field inside the sphere S_D of radius $D \geqslant D_{\min}$ concentric to $S_{\min}$ and are given by

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

The singular part is generated by the actual source and by the image sources 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 $S_{\max}$ (Figure 3; in the figure, $D = D_{\min}$, which is shown later to be the optimal choice for D; however, all derivations here are valid for any $D \geqslant D_{\min}$). Below the evaluation of the regular part $\psi_r(\mathbf{r})$ is considered.

Assume that a unit strength monopole source is located at a point s. Its potential $\psi(\mathbf{r}_{so})$ at any point

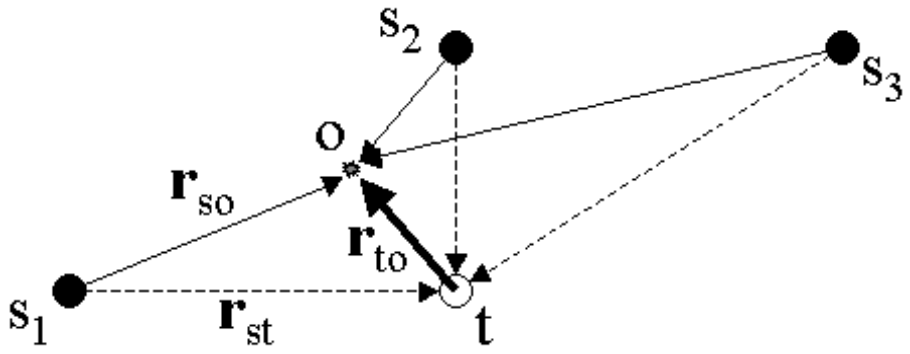

Fig. 4. An illustration of the translation operation. Monopole sources s_1, s_2, and s_3 create some potential at the evaluation point o. Instead of evaluating it directly as $\sum \psi(\mathbf{r}_{so})$, the potential of each source is expanded in terms of the functions centered about another point t (center of expansion); $\mathbf{r}_{st}$ denotes the translation vector and $\mathbf{r}_{to}$ is the vector from the center of expansion to the evaluation point. Expansion is done for all sources, and their combined potential $\sum \psi(\mathbf{r}_{so})$ at o can then be computed by evaluating, at point o, the potential of the multipole source located at t. Vectors $\mathbf{r}_{so}$ and $\mathbf{r}_{st}$ for sources 2 and 3 are not labeled for clarity.

o is given by the formula

$$\psi\left(\mathbf{r}_{so}\right) = \frac{e^{ik\mathbf{r}_{so}}}{4\pi|\mathbf{r}_{so}|} \tag{9}$$

(compare with equation (1)) and can be expanded as the product of a series of singular and regular basis functions of the Helmholtz equation centered about another point t using the following factorization identity ([38])

$$\psi\left(\mathbf{r}_{so}\right) = ik\sum_{n=0}^{\infty}\sum_{m=-n}^{m=n} S_n^{-m}\left(-\mathbf{r}_{st}\right) R_n^m\left(\mathbf{r}_{to}\right), \qquad r_{to} \leqslant r_{st}, \tag{10}$$

where $\mathbf{r}_{so}$ denotes the vector from the monopole source location s to the evaluation point o, $\mathbf{r}_{st}$ denotes a *translation vector* from the original monopole source location s to the center of expansion t, $\mathbf{r}_{to}$ denotes a vector from the center of expansion t to the evaluation point o (Figure 4), and

$$\begin{aligned} R_n^m\left(\mathbf{r}\right) &= j_n(kr)Y_n^m(\theta,\varphi), \\ S_n^m\left(\mathbf{r}\right) &= h_n(kr)Y_n^m(\theta,\varphi), \\ Y_n^m\left(\theta,\varphi\right) &= (-1)^m\sqrt{\frac{2n+1}{4\pi}\frac{(n-|m|)!}{(n+|m|)!}}P_n^{|m|}(\cos\theta)e^{im\varphi}. \end{aligned} \tag{11}$$

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

Next, all the image sources comprising the $\psi_r(\mathbf{r})$ in (8) are expanded in terms of the singular and regular basis functions of the Helmholtz equation centered at the origin of the coordinate system using equation (10). For each individual image source, the point s in equation (10) corresponds to its location

and the point t – to the origin (center of the room). Equation (10) is then substituted into the sum for the regular part of the potential field $\psi_r(\mathbf{r})$ in (8):

$$\psi_r(\mathbf{r}) = \sum_{D<r_q<R_{\max}} \left[ik \sum_{n=0}^{\infty} \sum_{m=-n}^{m=n} Q_q S_n^{-m}(\mathbf{r}_q) R_n^m(\mathbf{r}) \right] \tag{12}$$

and it is possible to regroup the summation to form the coefficients C_n^m (which do not depend of the evaluation point $\mathbf{r}$) as follows:

$$\psi_r(\mathbf{r}) = ik \sum_{n=0}^{\infty} \sum_{m=-n}^{m=n} C_n^m R_n^m(\mathbf{r}), \tag{13}$$

$$C_n^m = \sum_{D<r_q<R_{\max}} Q_q S_n^{-m}(\mathbf{r}_q), \tag{14}$$

where $\mathbf{r}$ is now just a vector from the origin to the evaluation point and $\mathbf{r}_q$ is a vector from the origin to the position of qth image source. The coefficients C_n^m do not depend on the evaluation point and can be precomputed for given image source locations (which are determined by simple geometric computations illustrated by Figures 1 and 2 in the same manner as in the original Allen-Berkley algorithm). In fact, the process of formation of these coefficients can be thought of as the process of collecting all the image sources constituting the regular part of the potential in equation (7) into a single (but rather complex) multipole source located at the center of the room. The potential generated by this multipole source inside the room matches exactly the sum of the potential generated by all collected image sources. As there exist tens of thousands image sources even in the simulation of as short as 100 ms RIR, a very significant speed-up of the algorithm can be achieved.

Equation (13) is exact. These series are very quickly convergent, and the outer summation can be truncated at $p+1$ terms:

$$\psi_r(\mathbf{r}) = ik \sum_{n=0}^{p} \sum_{m=-n}^{m=n} C_n^m R_n^m(\mathbf{r}). \tag{15}$$

The parameter p is known as the *truncation number*. Evaluation of $\psi_r(\mathbf{r})$ at any given point $\mathbf{r}$ thereby requires a summation of $(p+1)^2$ terms. This must be compared with the summation over N terms due to the image sources with $D < r_q < R_{\max}$ if equation (8) were used directly. Note that p only depends on the desired accuracy of the summation, and $(p+1)^2$ can be orders of magnitude smaller than N. Of course, it is still necessary to compute C_n^m first, which does require processing of N image sources; however, if $\psi_r(\mathbf{r})$ is to be computed at M evaluation points, the one-time cost of computing C_n^m is amortized over them, resulting in overall complexity decrease.

As the coefficients C_n^m do not depend on $\mathbf{r}$, the evaluation of the potential $\psi(\mathbf{r}; k)$ at M points is done in two steps. In the first step, C_n^m are computed using equation (14). In the second step, $\psi_r(\mathbf{r}; k)$ at M points is computed using the set of C_n^m coefficients produced in the first step, and $\psi_s(\mathbf{r}; k)$ is computed directly

($\psi_s(\mathbf{r};k)$ is generated by the actual source and by a few image sources inside S_D, so the computation time for $\psi_s(\mathbf{r};k)$ is negligible). $\psi_r(\mathbf{r};k)$ and $\psi_s(\mathbf{r};k)$ are then summed up to produce $\psi(\mathbf{r};k)$.

IV. Error Evaluation

Direct numerical implementation of (15) 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 the series can be truncated. 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 $R_n^m(\mathbf{r})$ in equation (11) can be found using asymptotic expressions for Bessel functions from Abramowitz & Stegun [43]

$$S_n^{-m}(\mathbf{r}_q) \sim -\frac{i}{kr_q}\sqrt{\frac{2}{e}}\left(\frac{2n+1}{ekr_q}\right)^n Y_n^{-m}(\theta_q,\varphi_q), \tag{16}$$
$$R_n^m(\mathbf{r}) \sim \frac{1}{2n+1}\sqrt{\frac{e}{2}}\left(\frac{ekr}{2n+1}\right)^n Y_n^m(\theta,\varphi).$$

These expressions allow one to determine the order of truncation [38]. Let us define

$$p_k(r) = \frac{ekr-1}{2}. \tag{17}$$

This has the following meaning: for given r_q and $n > p_k(r_q)$ 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 (16)

$$S_n^{-m}(\mathbf{r}_q)\,R_n^m(\mathbf{r}) = \left(\frac{r}{r_q}\right)^n \frac{Y_n^{-m}(\theta_q,\varphi_q)Y_n^m(\theta,\varphi)}{ikr_q(2n+1)}. \tag{18}$$

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, the optimal selection of truncation number as

$$p = \mu p_k(D), \tag{19}$$

$\mu = 1$, provides both stable computation of the series for the entire domain and the following evaluation of the truncation error for any point inside the room

$$\epsilon_p = O\left(\frac{1}{(2p+1)^2}\left(\frac{D_{\min}}{D}\right)^p\right). \tag{20}$$

The obtained result for the optimal value of the truncation number p is roughly consistent with the estimations presented earlier in the literature. For example, in [44] rough upper and lower bounds are established on the truncation number: first, the truncation number should be larger that $2kD$ for the multipole expansion to be valid (i.e., to produce the correct results) in the region of radius D; second,

the truncation number should not be much larger than $\min kr_{st}$, where r_{st} is the length of the translation vector $\mathbf{r}_{st}$ in equation (10) (in other words, if the nearest image source is located at range r^* and thus the shortest translation is done by distance r^*, then the p should not be much large than kr^*; otherwise, large-magnitude oscillations in the spherical Hankel functions cause loss of accuracy). Two important conclusions follow from these bounds:

- If the nearest image source is located at the distance which is comparable to the room radius D (which is usually true, as can be observed from Figure 3), the upper and lower bounds roughly coincide and approximately agree with equation (19).
- If D is reduced, p can also be reduced.

The second conclusion is important for the room subdivision technique described in Section 6.

V. Computational complexity

In this section, the computation complexity of the Allen-Berkley algorithm and the proposed multipole algorithm is considered. The complexity is first expressed in terms of the number of sources N and the number of evaluation points M. In the Allen-Berkley algorithm, computation of the field potential calls for the repeated evaluation of the potential for each source-receiver pair, resulting in $O(NM)$ operations. In contrast, when the potential is computed using the multipole algorithm, the first algorithm step – the computation of C_n^m – takes $O(N)$ time because the sum is taken over N image sources, and the second step takes $O(M)$ time because the truncated series (15) is computed once for each receiver point. The constant in the big-O notation depends on the truncation number p (i.e., on the desired accuracy). The total evaluation requires only $O(N + M)$ operations. The complexity decrease is more significant when M is large.

The regular structure of the image source lattice also permits evaluation of the computational complexity in terms of the simulated reverberation length $t_{\max}$. Note first that $R_{\max}$ is roughly proportional to $t_{\max}$, and the number of image sources in the $S_{\max}$ grows as $R_{\max}^3$ (because the problem is three-dimensional). The time-domain Allen-Berkley algorithm performs a summation of impulse responses of all image sources, and therefore its running time for M evaluation points is simply $O(Mt_{\max}^3)$.

The Allen-Berkley algorithm can also be implemented straightforwardly in the frequency domain by computing, in $O(t_{\max}^3)$ time (per receiver point), the field potential at each frequency presented in the RTF and then taking inverse Fourier transform to obtain RIR. The number of frequencies then is proportional to the length of RIR, and the running time (per one receiver point) of the frequency-domain Allen-Berkley algorithm is $O(t_{\max}^4)$. This running time is denoted T_{AB}. If the field is to be evaluated at M receivers, the total running time is $MT_{AB} = O(Mt_{\max}^4)$.

A faster hybrid-domain implementation is possible. It is based on computing, in the frequency domain, the spectrum for each image source using a limited number of frequency bands (e.g., 128 bands), obtaining the 256-point IR for that image source by IFFT, and inserting IRs for all image sources into the final RIR at appropriate time delays (which are dictated by the range of each image source) [45]. Because the number of the frequency bands is fixed, the total computation time grows only as $T'_{AB} = O(t^3_{\max})$. However, it should be noted here that the constant hidden in the big-O notation is quite large here because the computation is done at a fixed but pretty large (128) number of frequencies, and because the 256-point IFFT must be done for each image source. In the straightforward frequency-domain implementation the length of IFFT is the same as the length of RIR, but that IFFT is done only once after all image source contributions are assembled in the frequency domain. It will be shown in Section 8 that because of the reasons mentioned, the difference between T_{AB} and T'_{AB} is not as dramatic as it can be expected from $O(t^3_{\max})$ and $O(t^4_{\max})$ bounds. Similar technique with IIR (instead of FIR) filter is used in [21].

In contrast, the running time of the proposed algorithm consists of two parts, T_{MP1} and T_{MP2}. The first part T_{MP1} (the multipole source setup time) represents the computational time necessary to translate all the image sources to the origin; this procedure is done only once, no matter how many evaluation points are there in the room. The second part T_{MP2} represents the actual time to compute the reverberant field potential for one evaluation point.

Consider first the complexity for the case where the offline computation must be performed at multiple locations. If there are M receivers, the total running time is $T_{MP1} + MT_{MP2}$. As the first step of the multipole algorithm sums up all sources in $S_{\max}$ for all the frequencies in the RTF, T_{MP1} is $O(t^4_{\max})$. The second step however no longer involves processing all the image sources but rather computes the value of the multipole expansion for all frequencies so that T_{MP2} is $O(t_{\max})$. The total running time of the multipole algorithm is then $O(t^4_{\max} + Mt_{\max})$. It is evident that when M is sufficiently large, the multipole algorithm is faster than the frequency-domain and the time-domain Allen-Berkley algorithms.

In the second situation the evaluation of the RTF is performed online at a dynamically changing evaluation point. In this situation it is possible to neglect the running time of the first step of the multipole algorithm (for example, when the set of C_n^m is already precomputed (e.g., offline) for a given room and all that matters is the speed of RTF computation for an updated receiver position), and the running time (per evaluation point) of the time-domain (frequency-domain, hybrid-domain) Allen-Berkley algorithm grows as $t^3_{\max}$ ($t^4_{\max}$, $t^3_{\max}$, respectively), whereas the running time (per evaluation point) of the multipole method grows only as $t_{\max}$.

However, the performance evaluation results (presented in Section 8) reveal that for reasonable values of $t_{\max}$ and M the straightforward implementation of the multipole algorithm is significantly slower than

time-domain Allen-Berkley algorithm because of the computations of the special function values involved in the multipole expansion (10). To address this deficiency, a room subdivision method for improving the multipole algorithm performance is developed and is described in the next section. It should also be noted that if the simulation task necessitates the use of the frequency-domain algorithm, even the straightforward implementation of the proposed multipole algorithm is feasible and has an advantage in speed (also demonstrated in Section 8) over the hybrid-domain Allen-Berkley algorithm for commonly encountered values of $t_{\max}$ and M.

VI. Room Subdivision Technique

Due to the computational precision issues, it is problematic to raise the truncation number beyond a certain mark on a common hardware. It also may be desirable to speed up RIR computation to compete successfully with time-domain Allen-Berkley algorithm. The described algorithm can be modified as follows to decrease the required truncation number and to decrease T_{MP2}, at the expense of increasing T_{MP1}. The idea of the proposed modification is to split the room into several imaginary subrooms and create, in each subroom, a separate multipole source that is still built from *all* image sources but is used to evaluate the potential only for those evaluation points that belong to that subroom. Because the subrooms are smaller, the radius of validity of these multipole sources can be smaller. In this way, $D_{\min}$ and therefore p can be decreased.

Assume that each side of the room is divided into K pieces, creating a total of K^3 subrooms. Because subroom boundaries are imaginary, the image source lattice formation process (Figure 3) does not change. Then, for each subroom, all the image sources comprising the regular part of the potential are expanded in terms of the multipole source located at the subroom center, creating K^3 multipole sources, one per subroom. Each multipole source is now a local representation of all image sources (by "local", it is meant that the representation is valid only in the corresponding subroom). When a RTF is to be evaluated at some point, the subroom to which that point belongs is first found, and then the RTF is obtained by computing the potential of that subroom multipole source at the evaluation point and adding the singular part of the potential computed directly as before. In other words, each subroom created by subdivision is treated by the algorithm exactly in the same manner as the original room is treated in no subdivision case.

There are two advantages in using the room subdivision technique. Firstly, when the room is divided into K^3 subrooms, $D_{\min}$ gets divided by K. Therefore p also gets divided by K. This allows for computing the RTF at higher frequencies without hitting the machine precision limits (described in Section 8). Secondly, the number of terms in the multipole expansion and therefore T_{MP2}, as well as the time to compute C_n^m for one subroom, gets divided by K^2, allowing for faster RTF update in response to the source/receiver

motion (having precomputed all sets of C_n^m). Because there are total of K^3 subrooms, the T_{MP1} increases by a factor of K. However, this part of the calculation can be performed in advance for online applications. The room subdivision technique can therefore greatly increase the algorithm performance in situations where the sets of C_n^m can be precomputed beforehand so that T_{MP1} can be ignored and only T_{MP2} is of importance, and with appropriate subdivision the multipole algorithm can produce RTF faster than not only frequency-domain but also time-domain Allen-Berkley algorithm.

Room subdivision is also useful if the task is offline (so that T_{MP1} can not be ignored) and the number of evaluation points is large. Consider the case of computing the RIR at M points (e.g., for offline testing of the source localization or speech enhancement algorithm) so that the total computation time is $T_{MP} = T_{MP1} + MT_{MP2}$. As the first term increases and the second term decreases as K increase, it is possible to minimize T_{MP} by choosing K appropriately. Note that the first term grows proportionally to K and the second term grows proportionally to K^{-2}. Assume that the total number of sources is N; then, T_{MP} can be written as $NK + MK^{-2}$ (up to the scale factor). T_{MP} is minimized when its derivative with respect to K is equal to zero, and that derivative is $N - 2MK^{-3}$. Then, the optimal $K = (2M/N)^{1/3}$. It follows that for offline applications room subdivision is justified only when the number of evaluation points is comparable to the number of sources, which is usually not the case for RIR computation due to the large number of image sources.

VII. Results

To check the accuracy of the proposed algorithm, its implementation in the C programming language was developed and tested against the time-domain and against the frequency-domain implementations of the Allen-Berkley algorithm. Some representative results are shown below.

The first comparison is performed between the values of the potential obtained at a particular frequency using the exact equation (4) and using the approximation series (15), and the error is analyzed for various frequencies and for various truncation numbers p. The computations are carried out for the room ($l = 2.5$ m, $w = 2.5$ m, $h = 2.0$ m, $D_{\min} = 2.03$ m) with the reflection coefficients equal to 0.9 for room walls and 0.7 for the floor and the ceiling. The monopole source is placed at a random position inside the room. The radius $R_{\max} = 40$ m corresponds to $t_{\max} \approx 100$ms. In this particular configuration, approximately 21500 image sources are located within the sphere of radius $R_{\max}$.

To understand the behavior of the approximation error, M evaluation points are then distributed randomly over the room, and the field potential at these points is evaluated for a particular frequency and for varying truncation numbers, first by applying frequency-domain Allen-Berkley algorithm summarized by equations (4) and (5) (i.e., by summing Green's function of the actual source and all the image sources), and then by applying the developed method with $D = D_{\min}$. The root mean square approximation error

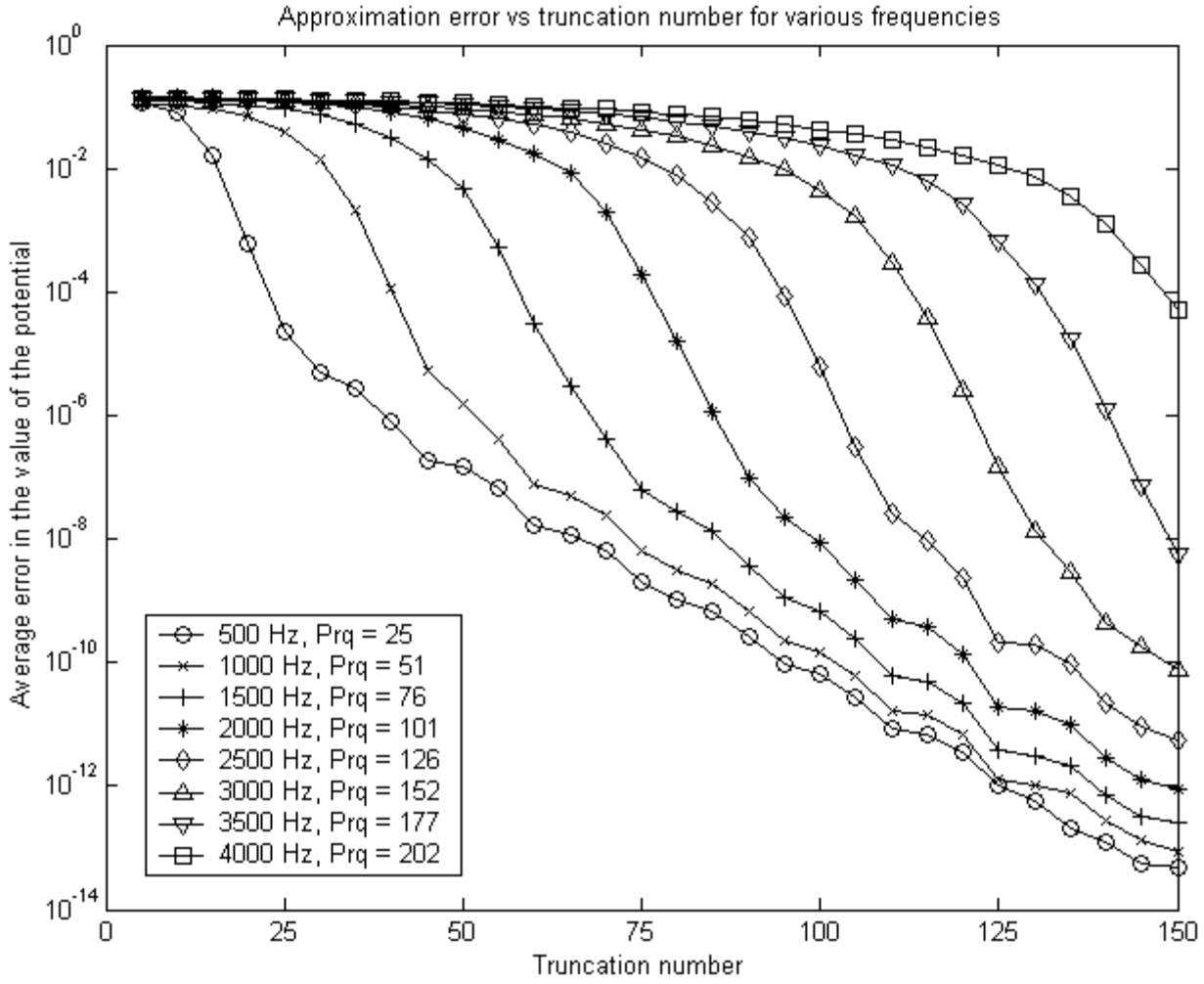

Fig. 5. Approximation error vs truncation number for various frequencies. For each frequency, the optimal truncation number Prq computed using equation (19) with μ=1 is shown in the legend.

is shown in Figure 5 for 8 frequencies and for the truncation number varying from 5 to 150. Figure 5 legend also provides the optimal truncation number (computed using equation (19) with $\mu = 1$) for each frequency. It can be seen from the plot that increasing μ beyond 1 does not have detrimental effect on computation accuracy (however, it does slow down computations as more coefficients are to be computed). It can also be seen that the error is indeed negligible for $\mu = 1$, and an approximation of reasonable quality can be obtained for lower μ. For example, the average approximation error is roughly 10%, 1%, 0.1%, and 0.01% for $\mu = 1/2$, $2/3$, $3/4$, and $4/5$, respectively. As it is computationally expensive to compute multipole expansion coefficients with high truncation numbers, in practical applications the compromise can be achieved by selecting the value of $\mu \leq 1$ depending on the desired accuracy, computational power available, and machine precision.

The computation of the actual RIR is then performed and RIRs obtained with the traditional time-domain

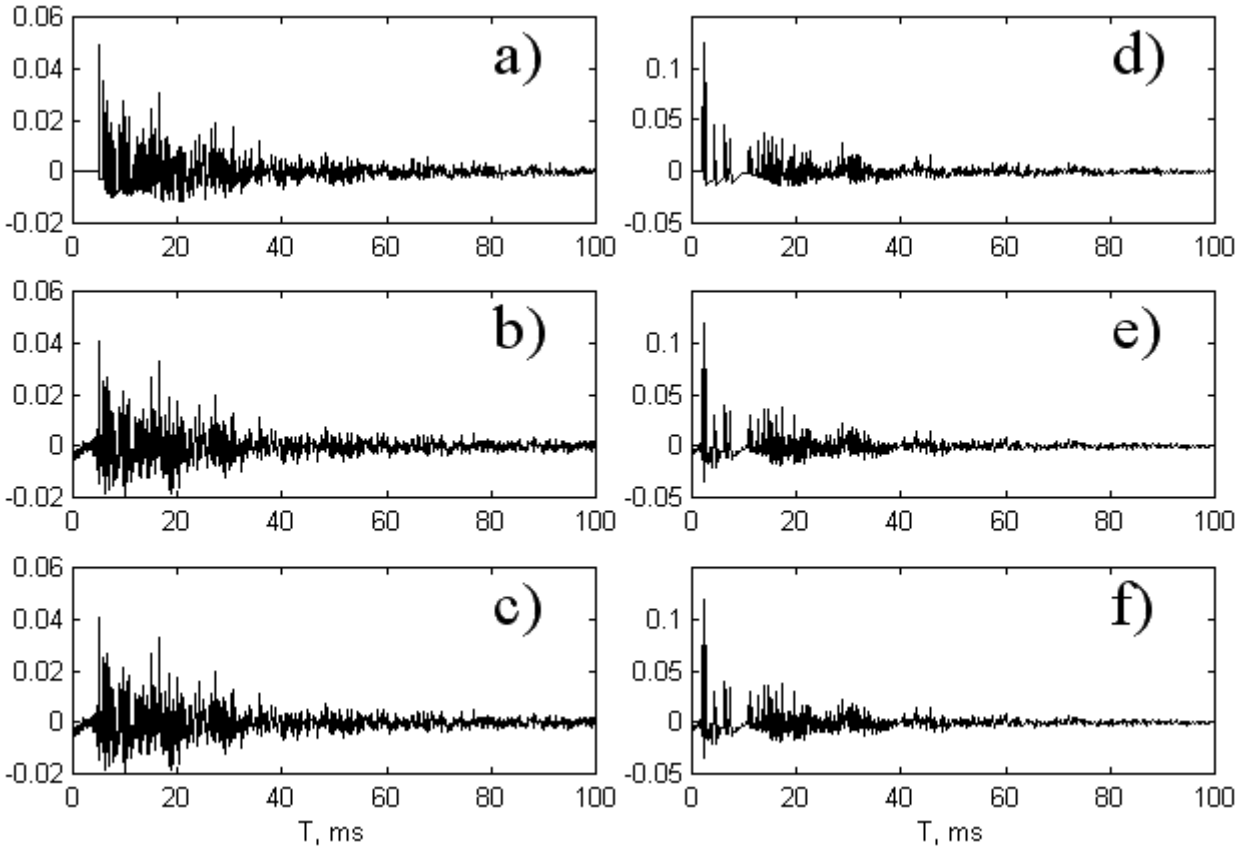

Fig. 6. a) RIR computed using time-domain Allen-Berkley algorithm for some source-receiver position pair. b) RIR computed using frequency-domain Allen-Berkley algorithm for the same source-receiver position pair. c) RIR computed using the proposed algorithm for the same source and receiver positions. d)-f) Same as a)-c) for another pair of source and receiver positions.

Allen-Berkley method, frequency-domain Allen-Berkley method, and the proposed multipole method are compared. The method for computing the RIR using the frequency-domain computation is as follows. Assume that the RIR is sampled at the sampling frequency f_d, and there are total of N_d points in RIR. The RTF then is simply a discrete Fourier transform of RIR and has a total of N_d points corresponding to the $N_d/2+1$ discrete frequencies $0, f_d/N_d, 2f_d/N_d, ..., f_d/2$. For each of these, excluding the zero frequency, the complex value of the potential at the evaluation point is computed using either method, and then the inverse discrete Fourier transform is applied to generate the RIR.

Figure 6 shows comparison between three algorithms for two randomly selected evaluation points. The discretization frequency here is 8 kHz, and $\mu = 3/4$ based on the error analysis shown above. For all three algorithms, time-domain high-pass filtering (as describes in [11]) with cutoff of 0.01 sampling frequency is applied to compensate for non-physical response at low frequencies due to no-flux boundary conditions. (Obviously, for frequency-domain methods filtering can be performed directly in the frequency domain before doing the inverse Fourier transform; this is not done here to preserve fairness of comparison). It can be seen that not much difference can be observed in the output of all three algorithms, and the

results produced by both frequency-domain algorithms are nearly identical. Some differences in peak heights can be seen between the time-domain and the frequency-domain Allen-Berkley algorithms. This difference is due to the fact that the time-domain algorithm rounds the arrival time to the nearest sample and places a single peak at that sample, whereas the frequency-domain algorithm is not obligated to do that, distributing the echo over several samples as necessary to preserve its exact arrival time. This effect was pointed out and handled by assignment of the low-passed impulse (instead of just shifted impulse) to each individual echo [13]; in the proposed algorithm, this problem is taken care of automatically. There is also a nonphysical artifact in the outputs of the frequency-domain and of the multipole algorithm in that the RIR is not zero before the direct sound arrival due to the wrap-around continuity of the inverse Fourier transform. This can be dealt with simply by zeroing the part of RIR before the direct arrival time, which is computed from the source and the receiver positions.

VIII. Performance and Implementation Considerations

As computational speed was the primary reason for the development of the proposed algorithm, the running time measurements are done in several simulation runs for variable modeled reverberation length. The simulated room was not changed from the setup described in the previous section. The discretization frequency of 8 kHz and μ equal to $3/4$ also did not change. Thus, for the maximum simulation frequency of 4 kHz $p = 151$, which is close to the maximum possible truncation number given the range of the floating point numbers allowable in the C compiler on PC-compatible architecture.

The running time of the frequency-domain Allen-Berkley algorithm, which essentially performs summation of the Green's functions corresponding to individual image sources evaluated at the receiver point, is linear (per one receiver point) in the number of the image sources. In contrast, the running time of the multipole algorithm consists of a fixed part T_{MP1} that is linear in the number of the image sources and variable part MT_{MP2} that is linear in the number of receivers and independent of the number of image sources. If $T_{MP2} < T_{AB}$, then the multipole algorithm has a distinct advantage over the summation of Green's functions because it becomes possible to compute the RTF faster per evaluation point (which is definitely of use in a real-time application such as MFA beamforming), albeit perhaps at a significant pre-computation expense incurred in the multipole source setup. Furthermore, even in the case of off-line computations, it might be necessary to compute the RTF at M points, and if $T_{MP2} < T_{AB}$, then at some M_0 (the break-even point) the total computation time of the multipole algorithm $T_{MP1} + M_0 T_{MP2}$ will be equal to $M_0 T_{AB}$, and for any $M > M_0$ it will be advantageous to use the multipole algorithm despite the lengthy first step.

Table I shows some experimental results for various reverberation length. It can be seen from the table that the multipole algorithm is slower per point for 50 ms reverberation, but becomes faster per point

$R_{\max}, m$	$t_{\max}, ms$	N	T_{AB}, s	T'_{AB}, s	T_{MP1}, s	T_{MP2}, s	M_0
20	~50	2680	0.37	0.25	1775	0.59	N/A
30	~75	9054	1.90	0.89	8551	0.90	8551
40	~100	21438	5.98	2.03	26714	1.21	5600
50	~125	41910	14.36	3.78	64577	1.46	5005
60	~150	72368	30.68	6.34	137100	1.79	4746

TABLE I

RUNNING TIME (PER EVALUATION POINT) FOR THE FREQUENCY-DOMAIN ALLEN-BERKLEY (AB) ALGORITHM (T_{AB}), HYBRID-DOMAIN AB ALGORITHM (T'_{AB}), AND THE PROPOSED MULTIPOLE ALGORITHM (T_{MP2}) FOR VARIABLE RADIUS OF INTEREST $R_{\max}$ AND REVERBERATION LENGTH $t_{\max}$. T_{MP1} IS THE TIME SPENT BY THE MULTIPOLE ALGORITHM IN THE PRE-CALCULATION PHASE, WHICH IS INDEPENDENT OF THE EVALUATION POINT LOCATION. THE TOTAL RUNNING TIME OF THE MULTIPOLE ALGORITHM IS $T_{MP1} + MT_{MP2}$, WHERE M IS THE NUMBER OF EVALUATION POINTS.

for longer reverberation time, with the break-even point M_0 steady decreasing as the reverberation length increases. It can also be noted that, as predicted in Section 5, T_{AB} grows as $t^4_{\max}$ whereas T_{MP2} grows only as $t_{\max}$. The multipole setup time T_{MP1} grows at the same rate as T_{AB}.

For reference, the running time T'_{AB} of the hybrid-domain Allen-Berkley algorithm is also presented in the Table I. The number of frequency bands in these simulations is equal to 32 (so that for each image source its 64-point IR is computed). The hybrid implementation is clearly faster than the straightforward frequency-domain implementation; however, T_{MP2} still grows slower than T'_{AB}, which can be expected from big-O bounds.

Still, the running time of the multipole algorithm is large in comparison with the time-domain Allen-Berkley algorithm (for example, it takes only 16.0ms to compute the room RIR using time-domain Allen-Berkley algorithm for $R_{\max} = 60$ m) because the time-domain algorithm has an additional advantage of not repeating computations at many frequencies (at a disadvantage of being unable to account for frequency dependencies in sound interaction with surfaces and materials). It should be noted that the proposed algorithm accounts for the contribution of all image sources in only 3.0ms (per frequency) for $R_{\max} = 60m$, which compares favorably with 16.0 ms required to sum up all the image sources by Allen-Berkley algorithm. It can also be noted that the values of $R_{\max}$ for which the straightforward implementation (without use of the room subdivision technique) of the multipole algorithm becomes faster per evaluation point than the time-domain Allen-Berkley algorithm are not practical, and T_{MP1} for such large $R_{\max}$ would be extremely large.

The running times given are obtained with Microsoft Visual C++ 6.0 optimizing compiler with full optimization on a dual-CPU Intel Pentium Xeon 1.7 GHz computer. Computations of the spherical Bessel

functions and spherical harmonics are based heavily on the routines described in the book [46] and freely available on the World Wide Web at [47]. Implementation of the algorithm for high truncation numbers becomes problematic because of the machine precision issues and limited range of the numbers that can be stored in double-precision variables in C. For example, the value of the spherical harmonic $Y_n^m(\theta, \varphi)$ given by equation (11) is a product of an associated Legendre polynomial (which has very large value for large m and n) and the square root in equation (11) (which has very small value for large m and n). The value of the square root is precomputed for all possible combinations of m and n up to the maximum possible truncation number using arbitrary precision arithmetic package to avoid expensive recomputations of the factorials of large numbers and to prevent overflow in the factorial computation. The minimum number that can be stored in a double precision variable on an Intel architecture is approximately $2.2 \cdot 10^{-308}$ (all smaller values are represented as zero), and the value of the root in equation (11) underflows at $m = n = 155$. As a results, the computed value of some spherical harmonics (specifically, the ones with high m and n) becomes zero if $p > 155$, even given that the value of the associated Legendre polynomial is about 10^{302} for $m = n = 155$ and after multiplication the result has close to unit magnitude. The values of the associated Legendre polynomial and of the imaginary part of the spherical Hankel functions (computed from the values and from the derivatives of the spherical Bessel functions) would also soon present overflow problem if p were to be further increased. Because of these issues, the maximum possible truncation number for the straightforward implementation is 155, with the quality of approximation degrading beyond that point due to zeroing out and/or overflowing of more and more of the basis functions in the multipole expansion, and with $p = 155$ reasonable-quality approximation is obtained for the frequencies up to 6 kHz. Computations with higher p can be carried out by implementing the algorithm with the help of an arbitrary precision arithmetic software package (which is likely to be extremely slow), on a different architecture computer that has wider range of double precision numbers (such as one based on the Motorola 680X0 or Sun SPARC processors, which have range of double precision numbers from 10^{-4932} to 10^{4932}), or even on a specialized hardware.

An alternative method of raising the high frequency limit is to use the room subdivision technique described in Section 6, which allows p to be decreased. Preliminary tests using the simulated setup described above with $K = 2$ and $R_{\max} = 40$ m show that results of the same quality are obtained at $p = 75$ at the frequency of 4 kHz. Another advantage of the room subdivision method is that T_{MP2} gets decreased (at the expense of increasing T_{MP1}). Experiments done with $K = 2$ show that T_{MP1} indeed grows from 26514 seconds to 51223 seconds (approximately by a factor of K) and T_{MP2} falls from 1.21 seconds to 0.29 seconds (approximately by a factor of K^2), confirming the theoretical predictions. Note that T_{MP2} now allows for near real-time operations. Additional subdivisions decrease T_{MP2} even further.

IX. Conclusions

An algorithm for fast frequency-domain computation of the room transfer function is proposed and tested. The algorithm is based on multipole expansion of the potentials generated by image sources. A comparison with the classical Allen-Berkley algorithm shows that the computations with similar accuracy can be performed at substantial speed-ups. 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 with computational complexity similar to the perception-oriented approximate modeling techniques such as those currently used in auralization.

References

[1] R. Duraiswami, N. A. Gumerov, D. N. Zotkin, L. S. Davis (2001). “Efficient evaluation of reverberant sound fields”, Proc. IEEE WASPAA 2001, New Paltz, NY, October 2001, pp. 203-206.

[2] W. C. Sabine (1900). “Reverberation”, originally published in 1900 and reprinted in Acoustics: Historical and Philosophical Development, ed. by R. D. Lindsay, published by Dowden, Hutchinson, and Ross, Stroudsburg, PA (1972).

[3] Y. Ando (1985). “Concert hall acoustics”, Springer-Verlag, Berlin, Germany.

[4] H. Kuttruff (1991). “Room acoustics (3rd edition)”, Elseiver Applied Science, Netherlands.

[5] W. M. Hartmann (1997). “Listening in a room and the precedence effect”, in Binaural and Spatial Hearing in Real and Virtual Environments, ed. by R. H. Gilkey and T. R. Anderson, Lawrence Erlbaum Associates, Mahwah, NJ, pp. 191-210.

[6] B. G. Shinn-Cunningham, J. G. Desloge, and N. Kopco (2001). “Empirical and modeled acoustic transfer functions in a simple room: effects of distance and direction”, Proc. IEEE WASPAA 2001, New Paltz, NY, pp. 419-423.

[7] J. L. Flanagan, A. C. Surendran, and E. E. Jan (1993). “Spatially selective sound capture for speech and audio processing”, Speech Communication, vol. 13, pp. 207-222.

[8] J. H. DiBiase, H. F. Silverman, and M. S. Brandstein (2001). “Robust localization in reverberant rooms”, in Microphone Arrays: Signal Processing Techniques and Applications, ed. by M. S. Brandstein and D. B. Ward, Springer-Verlag, Berlin, Germany, pp. 157-180.

[9] M. R. Schroeder (1962). “Natural sounding artificial reverberation”, J. Audio Eng. Soc., vol. 10, no. 3, pp. 219-223.

[10] J. A. Moorer (1979). “About this reverberation business”, Computer Music Journal, vol. 3, no. 2, pp. 13-28.

[11] J. B. Allen and D. A. Berkley (1979). “Image method for efficiently simulating small-room acoustics”, J. Acoust. Soc. Am., vol. 65, no. 5, pp. 943-950.

[12] J. Borish (1984). “Extension of the image model to arbitrary polyhedra”, J. Acoust. Soc. Am., vol. 75, no. 6, pp. 1827-1836.

[13] P. M. Peterson (1986). “Simulating the response of multipole microphones to a single acoustic source in a reverberant room”, J. Acoust. Soc. Am., vol. 80, no. 5, pp. 1527-1529.

[14] M. Kompis and H. Dillier (1993). “Simulating transfer functions in a reverberant room including source directivity and head-shadow effects”, J. Acoust. Soc. Am., vol. 93, no. 5, pp. 2779-2787.

[15] J. S. Suh and P. A. Nelson (1999). “Measurement of transient response of rooms and comparison with geometrical acoustic models”, J. Acoust. Soc. Am., vol. 105, no. 4, pp. 2304-2317.

[16] T. D. Abhayapala, R. A. Kennedy, and R. C. Williamson (1999). “Noise modeling for nearfield array optimization”, IEEE Signal Processing Letters, vol. 6, pp. 210-212.

[17] M. S. Brandstein (1999). “Time-delay estimation of reverberated speech exploiting harmonic structure”, J. Acoust. Soc. Am., vol. 105, no. 5, pp. 2914-2919.

[18] M. Omologo, P. Svaizer, and M. Matassoni (1998). “Environmental conditions and acoustic transduction in hands-free speech recognition”, Speech Communication, vol. 25, pp. 75-95.

[19] P. W. Shields and D. R. Campbell (1998). “Intelligibility improvements obtained by enhancement method applied to speech corrupted by noise and reverberation”, Speech Communication, vol. 25, pp. 165-175.

[20] T. Funkhouser, N. Tsingos, I. Carlbom, G. Elko, M. Sondhi, J. E. West, G. Pingali, P. Min, and A. Ngan (2004). “A beam tracing method for interactive architectural acoustics”, J. Acoust. Soc. Am., vol. 115, no. 2, pp. 739-756.

[21] L. Savioja, J. Huopaniemi, and T. Lokki (1999). “Creating interactive virtual acoustic environments”, J. Audio Eng. Soc., vol. 47, no. 9, pp. 675-705.

[22] N. Tsingos, T. Funkhouser, A. Ngan, and I. Carlbom (2001). “Modeling acoustics in virtual environments using the uniform theory of diffraction”, Proc. ACM SIGGRAPH 2001, Los Angeles, CA, pp. 545-552.

[23] L. Cremer and H. A. Muller (1978). “Principles and applications of room acoustics”, vol. 1, Applied Science Publishers, London, UK.

[24] D. R. Begault (1994). “3D sound for virtual reality and multimedia”, Academic Press Professional, Boston, MA.

[25] W. Ahnert and R. Feistel (1993). “EARS auralization software”, J. Audio Eng. Soc., vol. 41, no. 1, pp. 894-904.

[26] J.-M. Jot, L. Cerveau, and O. Warusfel (1997). “Analysis and synthesis of room reverberation based on a statistical time-frequency model”, Proc. 103th Audio Eng. Soc. Conv., preprint 4629.

[27] J. F. Abbott (1991). “The interaction of sound and shock waves with flexible porous materials”, Ph.D. thesis, Department of Physics, MIT, Cambridge, MA.

[28] American national standard method for the calculation of the absorption of sound by the atmosphere. ANSI SI.26-1978, American Institute of Physics (for Acoustical Society of America), New York, 1978.

[29] W. G. Gardner (1992). “The virtual acoustic room”, M.S. thesis, Department of Computer Science and Engineering, MIT, Cambridge, MA.

[30] N. Tsingos, I. Carlbom, G. Elko, R. Kubli, and T. Funkhouser (2002). “Validation acoustical simulations in the Bells Labs box”, IEEE Computer Graphics and Applications, July/August 2002, pp. 28-37.

[31] E.-E. Jan, P. Svaizer, and J. L. Flanagan (1995). “Matched-filter processing of microphone array for spatial volume selectivity”, Proc. IEEE ISCAS 1995, Seattle, WA, pp. 1460-1463.

[32] R. J. Renomeron, D. V. Rabinkin, J. C. French, and J. L. Flanagan (1997). “Small-scale matched filter array processing for spatially selective sound capture” (abstract), J. Acoust. Soc. Am, vol. 105, no. 5, p. 3208 (Proc. 143th JASA meeting, December 1997).

[33] D. Rabinkin, R. Renomeron, J. Flanagan, D. F. Macomber (1998). “Optimal truncation time for matched filter array processing”, Proc. IEEE ICASSP 1998, vol. 6, pp. 3629-3632, Seattle, WA.

[34] P. M. Morse and K. U. Ingard (1968). “Theoretical acoustics”, Princeton Univ. Press, NJ.

[35] V. Rokhlin (1983). “Rapid solution of integral equations of classical potential theory”, Comput. Phys., vol. 60, pp. 187-207.

[36] T. Sakama and Y. Yasuda (2002). “Fast multipole boundary element method for large-scale steady-state sound field analysis. Part I: Setup and validation”, Acustica/Acta Acustica, vol. 88, pp. 513-525.

[37] Y. Yasuda and T. Sakama (2003). “Fast multipole boundary element method for large-scale steady-state sound field analysis. Part II: Examination of numerical items”, Acustica/Acta Acustica, vol. 89, pp. 28-38.

[38] N. A. Gumerov and R. Duraiswami (2004). “Fast multipole methods for the Helmholtz equation in three dimensions”, Elsevier Science, Netherlands.

[39] M. Kleiner, B.-I. Dalenback, and P. Svensson (1993). “Auralization – An overview”, J. Audio Eng. Soc., vol. 41, no. 11, pp. 861-875.

[40] M. Vorlander (1995). “International round robin on room acoustical computer simulations”, Proc. 15th Intl. Congress on Acoustics, Trondheim, Norway, June 1995, pp. 689-692.

[41] I. Bork (2000). “A comparison of room simulation software – the 2nd round robin on room acoustical computer simulation”, Acustica/Acta Acustica, vol. 86, pp. 943-956.

[42] A. Krokstad, S. Strom, and S. Sorsdal (1968). “Calculating the acoustical room response by the use of a ray tracing technique”, Journal of Sound and Vibration, vol. 8, no. 1, pp. 118-125.

[43] M. Abramowitz and I. A. Stegun (1964). “Handbook of mathematical functions”, National Bureau of Standards, Washington, D.C.

[44] R. Coifman, V. Rokhlin, and S. Wandzura (1993). “The fast multipole method for the wave equation: a pedestrian prescription”, IEEE Antennas and Propagation Magazine, vol. 35, pp. 7-12.

[45] D. van Maercke and J. Martin (1993). “The prediction of echograms and impulse responses within the Epidaure software”, Applied Acoustics, vol. 38, pp. 93-114.

[46] S. Zhang and J. M. Jin (1996). “Computation of special functions”, John Wiley and Sons, New York.

[47] http://jin.ece.uiuc.edu/routines/routines.html.