Applied Acoustics **42** (1994) 313–332

Simulation of the Response of Multiple Microphones to a Moving Point Source

Benoit Champagne

INRS-Télécommunications, Université du Québec, 16 place du Commerce, Verdun, Québec, Canada, H3E 1H6

(Received June 1993; revised version received 2 September 1993; accepted 30 September 1993)

ABSTRACT

New methods are presented for the accurate and efficient simulation of the response of multiple microphones to a moving point source in a reverberant space. The development of these methods is based on an extension of the image model technique to moving sources. It is shown that source motion introduces shifts in the propagation delays associated with the images in the model. Even for a moderate source speed, these shifts can significantly alter the relative phase between the outputs of multiple microphones. The proposed methods incorporate accurate modeling of these effects and can be used reliably in the study of multi-microphone systems that are sensitive to intermicrophone phase errors. The methods, which differ in the way the time-varying propagation delays are recomputed, offer different advantages in terms of computational complexity and simulated waveform accuracy. Computer implementations of the methods based on a rectangular room model are described. Finally, simulation results are used to illustrate their comparative behavior.

1 INTRODUCTION

Recently, the use of multi-microphone systems for sound transduction under reverberant and noisy conditions has been proposed in several applications, including audio-conferencing,[1] hearing aids,[2] and mobile telephony.[3] By means of so-called beamforming algorithms, an array of spatially distributed microphones can be steered in the direction of an

Applied Acoustics 0003-682X/94/$07.00

active source, while attenuating interfering signals that are coming from unwanted directions. This results in considerable improvements in the perceived quality of the captured audio signal.

Computer generation of synthetic microphone signals plays a very important role in the study and design of multi-microphone systems. By using simulated data it is possible to investigate the comparative behavior of different beamforming algorithms under a wide range of controlled conditions. At the design level, simulated data can be used effectively to verify that a particular system (i.e., microphone configuration and processing algorithm) meets the given performance specifications.

Computer simulation of the response of a microphone to an acoustic source in a reverberant room was first considered by Allen and Berkley.[4] Using the image model technique, they developed a computer program for calculating the impulse response between a pair of (fixed) omnidirectional transmitting and receiving points in a rectangular room. The desired microphone response was produced by convolving the calculated impulse response with an anechoic version of the source signal.

Allen and Berkley's approach for computing the impulse response of a reverberant room, also known as the *shifted impulse method*, uses uniform sampling in the time domain and therefore introduces small errors in the computed echo arrival times. The magnitude of these errors depends on the sampling rate, which is typically chosen to be the Nyquist rate of the source signal in order to facilitate the convolution operation. These quantization errors, usually negligible in single-microphone applications, become critical when the method is applied to multi-microphone systems that are sensitive to intermicrophone phase errors. To overcome this difficulty, one possible approach is to increase the sampling rate in the shifted impulse method. Another possibility is to use the *low-pass impulse method* originally proposed by Peterson,[5] in which each received echo is represented by the samples of a low-pass filtered impulse at the correct arrival time.†

Both the shifted and low-pass impulse methods are based on the assumption of fixed source and microphone positions. In certain applica-

† The use of a higher sampling rate in the shifted impulse method is done at the expense of additional memory needed to store the discrete-time impulse response, but no extra computation is necessary. Moreover, by allowing the sampling period to be sufficiently small, the echo arrival times can be computed with an arbitrary precision (up to machine accuracy). At the end of the computation, the impulse response can be down-sampled to the Nyquist rate by using an appropriate low-pass filter. This approach has at least two advantages over the low-pass impulse method. Indeed, in the latter, each received echo is represented by calculated samples of a shifted low-pass filter function, so that extra computations are necessary. Moreover, to make the procedure tractable, only simple low-pass filter functions can be used, a constraint which is not required in the shifted impulse method.

tions, however, it is desirable to simulate the response of multiple microphones to a moving, rather than a fixed source. This occurs for example in the design of self-steering microphone arrays with capabilities to track a moving source (e.g., talker in audio-conferencing room) or in the study of noise fields produced by moving objects. In such cases, the above methods can be applied informally by computing a set of stationary room impulse responses corresponding to different positions along the source trajectory. However, as we shall see later, this approach is not computationally efficient and it can introduce important errors in the calculated echo arrival times if it is not implemented with care. Even for moderate source velocities, these errors can exceed the quantization errors typically introduced by the shifted impulse method at the Nyquist rate.

In this paper, we present three methods for accurately and efficiently simulating the response of multiple microphones to a moving point source in a reverberant space. Given the original source signal (e.g., speech), the trajectory of the source, the geometry and acoustic properties of the environment, and the microphone locations, the proposed methods can be used to generate accurate replicas of the microphone output signals. Because of their inherent accuracy, these synthetic signals can be used reliably in the investigation of multi-microphone systems. The proposed methods, which are based on an extension of the image model technique to moving sources, differ in the way the time-varying image structure is recomputed and offer different advantages in terms of computational complexity and waveform accuracy.

The paper is organized as follows. The extension of the image model technique to a moving source is considered in Section 2. The proposed simulation methods are described in Section 3, along with specific computer implementations based on a rectangular room model. The comparative performance of these methods is illustrated in Section 4 by means of simulation results. A summary and some additional discussions are presented in Section 5.

2 EXTENSION OF THE IMAGE MODEL TECHNIQUE

Consider a reverberant space bounded by plane reflecting surfaces and let a particular point in this space be the origin of a rectangular coordinate system $Oxyz$. Let $\mathbf{x} = (x, y, z)$ denote the position vector of a fixed omnidirectional point acoustic source. Similarly, let $\mathbf{x}' = (x', y', z')$ denote the position vector of an omnidirectional point receiver (the microphone) used to monitor the source signal. According to the image model tech-

nique,[6] the acoustic signal produced at position $\mathbf{x}'$ by an impulse excitation at position $\mathbf{x}$ and time $t = 0$ is given by

$$h(t) = \frac{1}{4\pi} \sum_r \frac{\beta_r}{|\mathbf{x}_r - \mathbf{x}'|} \delta(t - |\mathbf{x}_r - \mathbf{x}'|/c) \tag{1}$$

where r is the image index ($r = 0$ usually corresponds to the source itself), $\mathbf{x}_r$ is the position vector of the image indexed by r, β_r is the corresponding composite reflection coefficient, $\delta(.)$ is the Dirac delta function and c is the speed of sound in air.

Equation (1) is equivalent to adding the responses produced at $\mathbf{x}'$ by multiple image sources (possibly an infinity) located at positions $\mathbf{x}_r$. In this respect, the term $|\mathbf{x}_r - \mathbf{x}'|/c$ in the argument of the delta function represents the propagation delay from the rth image to the receiver, while the multiplicative factor $1/|\mathbf{x}_r - \mathbf{x}'|$ represents the attenuation for spherical wavefront propagation. The multiplicative factor β_r accounts for attenuation produced by successive reflections on the planar boundaries. Implicit in eqn (1) is the assumption of frequency-independent reflection coefficients. To reduce the amount of computation involved, it is further assumed in many applications that the reflection coefficients are independent of the angle of incidence. It is then possible to pre-compute and store the composite reflection coefficients β_r prior to the application of the image model technique.

For a rectangular room, the image distribution forms a three-dimensional rectangular lattice and can be calculated easily. Let the center of the room be the origin of the coordinate system $Oxyz$, with axes parallel to the walls, and let L_x, L_y and L_z denote the dimensions of the room along these axes. The images are indexed with an integer triplet $r = (r_1, r_2, r_3)$ and the summation in eqn (1) is over all possible values of the integers $r_i = 0, \pm 1, \cdots$. The position vector $\mathbf{x}_r$ and the composite reflection coefficient β_r are given by

$$\mathbf{x}_r = (r_1 L_x + (-1)^{r_1} x,\ r_2 L_y + (-1)^{r_2} y,\ r_3 L_z + (-1)^{r_3} z) \tag{2}$$

$$\beta_r = \beta_{x+}^{n_1^+} \beta_{x-}^{n_1^-} \beta_{y+}^{n_2^+} \beta_{y-}^{n_2^-} \beta_{z+}^{n_3^+} \beta_{z-}^{n_3^-} \tag{3}$$

In eqn (3), β_{x+} denotes the reflection coefficient of the wall intersecting the positive x-axis at a right angle, while n_1^+ denotes the number of reflections of the ray path from image r to the receiver on this wall. A similar notation is used for the other walls. The exponents $n_i^{\pm}$ are given by

$$n_i^+ = |\lceil r_i/2 \rceil|, \qquad n_i^- = |\lfloor r_i/2 \rfloor| \tag{4}$$

where $\lceil . \rceil$ is the smallest integer not less than its argument and $\lfloor . \rfloor$ is the largest integer not greater than its argument. At the expense of increased

computational complexity, the image model technique can be extended to geometries other than rectangular. Borish[7] and Lee and Lee[8] describe algorithms for calculating the image distribution in the case of an arbitrary polyhedral space. These algorithms can be used jointly with the simulation methods proposed in this paper.

Now, suppose that the source is moving and let $\mathbf{x}(t) = (x(t), y(t), z(t))$ denote its position vector at time t. (For simplicity, we shall assume that the receiver is fixed, although receiver motion could be included in the analysis.) To extend the image model technique to this situation, we follow a standard approach which consists of modeling the acoustic space as a time-invariant, linear physical system.[6]

Invoking the assumption of temporal invariance, it follows from eqn (1) that the signal produced at position $\mathbf{x}'$ by an impulse excitation at position $\mathbf{x}(u)$ and time u is

$$h(t,u) = \frac{1}{4\pi} \sum_r \frac{\beta_r}{|\mathbf{x}_r(u) - \mathbf{x}'|} \delta(t - u - |\mathbf{x}_r(u) - \mathbf{x}'|/c) \tag{5}$$

Let $a(t)$ denote the audio signal emitted by the moving source. This signal can be represented as a linear superposition of successive impulses emitted along the source trajectory described by $\mathbf{x}(u)$. Invoking the assumption of linearity, it follows that the signal produced at $\mathbf{x}'$ by the moving source can be written as the convolution integral

$$s(t) = \int_{-\infty}^{\infty} h(t,u)a(u)\, du \tag{6a}$$

$$= \frac{1}{4\pi} \sum_r \beta_r \int_{-\infty}^{\infty} \delta\left(t - u - \frac{g_r(u)}{c}\right) \frac{a(u)}{g_r(u)} du \tag{6b}$$

where the function

$$g_r(u) = |\mathbf{x}_r(u) - \mathbf{x}'| \tag{7}$$

gives the distance between the image r and the receiver at time u.

The evaluation of the integral in eqn (6b) requires an analysis of the zeros of the function $f(u; r, t) = t - u - g_r(u)/c$, where u is the free variable and the image index r and the time t are treated as parameters. It can be verified that if the source velocity is smaller than the speed of sound in air, i.e. if $|\dot{\mathbf{x}}_r(u)| < c$ for all u, where $\dot{\mathbf{x}}_r(u)$ is the time derivative of $\mathbf{x}_r(u)$, then the function $f(u; r, t)$ has a unique zero for every value of the parameters r and t. This zero is given by $u = t - \tau_r(t)$, where $\tau_r(t)$ is defined implicitly by the equation

$$\tau_r(t) = \frac{1}{c} g_r(t - \tau_r(t)) \tag{8}$$

The quantity $\tau_r(t)$ represents the travel time from image r to the microphone, for the signal component reaching the microphone at time t.

Having identified the unique zero of $f(u; r, t)$, the integral in eqn (6b) can be evaluated using well-known properties of the delta function.[9] The result is

$$s(t) = \frac{1}{4\pi c} \sum_r \frac{\beta_r[1 - \dot{\tau}_r(t)]}{\tau_r(t)} a(t - \tau_r(t)) \tag{9}$$

where $\dot{\tau}_r(t)$ is the time derivative of $\tau_r(t)$. The right-hand side of eqn (9) represents a time-variant linear filtering operation on the transmitted signal $a(t)$. It consists of a sum (possibly infinite) of time-delayed and scaled versions of $a(t)$, one for each image source in the model. However, due to the relative motion between the source and the receiver, the corresponding time delays and scaling factors are time-dependent.

In the case of a narrow-band audio signal $a(t)$, the time-varying nature of the delay $\tau_r(t)$ in the argument of $a(t - \tau_r(t))$ in eqn (9) causes a shift of the center frequency of the received echo, i.e., Doppler effect.[9] The sign and magnitude of this frequency shift will actually be different for different images. More generally, source motion introduces a shift in the propagation delay $\tau_r(t)$, relative to the delay that would be obtained with a fixed source at position $\mathbf{x} = \mathbf{x}(t)$. The value of this shift at time t is given by

$$\Delta_r(t) = \tau_r(t) - \hat{\tau}_r(t) \tag{10}$$

where

$$\hat{\tau}_r(t) = \frac{1}{c} g_r(t) \tag{11}$$

Numerical values for eqn (10) are calculated in Section 3 where a specific example is considered. In addition to frequency or time delay shifts, source motion alters the amplitude of the received echoes by a factor $[1 - \dot{\tau}_r(t)]$, when compared to spherical wavefront propagation. This factor is positive if the image r is moving towards the receiver; otherwise it is negative.

When the source position is fixed, the function $g_r(t)$ in eqn (7) is constant over time, the propagation delays $\tau_r(t)$ in eqn (8) are also constant, and eqn (9) reduces to a standard form.

3 IMPLEMENTATION ISSUES

As indicated in Section 1, both the shifted and low-pass impulse methods can be used informally to calculate the microphone response $s(t)$ in eqn (6a). This is achieved by first computing a set of stationary impulse

responses corresponding to different (fixed) positions along the source trajectory, say $\mathbf{x}(u_k)$ for $k = 0, \pm 1, \ldots$, and by partitioning the time axis into subintervals delimited by the times u_k. The convolution integral of eqn (6a) is then decomposed into a sum of partial convolutions over these subintervals. Finally, each partial convolution is approximated using the corresponding stationary impulse response. However, as explained below, this direct approach is not computationally efficient and can introduce important errors in the computed echo arrival times if it is not implemented carefully.

The above approach is equivalent to assuming that the source moves in a discontinuous manner, jumping from one position to the next in a very short amount of time and remaining at the new position for some time. Clearly, this approach introduces discontinuities in the computed echo arrival times and does not accurately reproduce the Doppler effect. To minimize the resulting errors in the simulated microphone response $s(t)$, the stationary impulse responses must therefore be calculated on a sufficiently dense grid of points on the source trajectory. This is particularly important here since no interpolation mechanism of the impulse response is used to smooth the transition between adjacent points on the grid.

Another problem can be described as follows. For a given grid of positions along the source trajectory, the length of the subintervals of integration in the partial convolutions is inversely proportional to the source velocity. In the case of a moving source, due to the long duration of the impulse response in a reverberant room, it is therefore necessary to use several stationary impulse responses based on different source positions in the convolution of eqn (6a). If the convolution sum is not decomposed in this manner and a single stationary impulse response based on the current source position is used, important errors in the computed echo arrival times of distant images will be introduced (for additional explanations, see Section 3.3 below).

Hence, even for moderate source velocities, this direct approach poses serious computational problems. In this section, we develop new methods for the accurate simulation of the microphone response $s(t)$ that are based on the representation of eqn (9) instead of eqn (6a). The use of eqn (9) enables us to deal more efficiently with the computational difficulties posed by source motion.

Equation (9) provides a closed-form, continuous-time representation for the response of a microphone to a moving point source. However, its implementation on a digital computer raises several issues, namely:

(1) In most applications, the number of images included in the summation over r is theoretically infinite.

(2) In computer simulations, only a sampled (discrete-time) version of the transmitted signal $a(t)$ is available. This poses a problem since the delays $\tau_r(t)$ are generally not integer multiples of the sampling period.
(3) Except for trivial cases, the calculation of the delays $\tau_r(t)$ in eqn (8) involves the solution of a non-linear equation.
(4) The delays $\tau_r(t)$ vary with time and must be recomputed on a regular basis.

These issues are addressed individually in Sections 3.1 to 3.4 below, leading to three distinct methods for the evaluation of eqn (9) on a digital computer. Computer implementations of these methods based on a rectangular room model are described in Section 3.5. Since the first issue can be handled exactly as in the fixed source case, it is considered only briefly.

3.1 Truncation of image model

As indicated above, the summation in eqn (9) usually contains an infinity of terms. However, because the amplitude of a received echo decreases as the distance from its origin (i.e. the image source) and the number of its successive reflections increase, only a finite number of images need to be included. In the case of a reverberant enclosure, two simple criteria can be used to truncate the summation. In the first criterion, only those images whose order (the order of an image is defined as the number of successive reflections corresponding to that image[8]) does not exceed a preset value are included in the sum. In the second criterion, only those images within a specific distance from the origin are included. A typical choice for this distance is cT_R, where T_R is the reverberation time of the enclosure, calculated according to Sabine's formula.[6]

3.2 Interpolation of sampled signal

The time delays $\tau_r(t)$ in eqn (8) are generally not integer multiples of the sampling interval T_s at which the source signal $a(t)$ is available in a given application. Yet, it is critical to retain the fractional part of the delays when considering multi-microphone systems that are sensitive to intermicrophone phase. In the case of a fixed source (where $\tau_r(t) = \tau_r$, a constant), this problem can be overcome either by increasing the sampling rate in the shifted impulse method or by using the low-pass impulse method. In both cases, digital convolution of the calculated impulse response with the signal samples $a(nT_s)$ is equivalent

to performing a band-limited interpolation on these samples in order to obtain the delayed signals $a(nT_s - \tau_r)$ corresponding to the individual echoes.

In the proposed digital implementations of eqn (9), the delayed signal $a(nT_s - \tau_r(t))$ is obtained by performing a band-limited interpolation directly on the signal samples $a(nT_s)$. The interpolation is done in two stages following a conventional procedure:[10] $L - 1$ zeros are padded between each signal sample, where L is the upsampling factor, and the padded sequence is filtered with a low-pass interpolation filter. A user-defined symmetric finite impulse response (FIR) filter can be used for this purpose.

A circular buffer of length K can be used to store the most recent interpolated signal values, i.e. $a(nT_s)$, $a(nT_s - T_s/L)$,..., $a(nT_s - (K - 1)T_s/L)$. (In practise, the length of the circular buffer can be set to $K = L\tau_{max}/(K = \tau_{max}/T_s)T_s$, where τ_{max} is an upper bound on the propagation delays $\tau_r(t)$.) After the delays $\tau_r(t)$ and their time derivatives $\dot{\tau}_r(t)$ have been calculated with sufficient accuracy (see Sections 3.3 and 3.4 below), this buffer is read and the microphone output $s(nT_s)$ is calculated according to eqn (9). In this respect, it has been found that one further step of linear interpolation between consecutive elements of the buffer significantly improves the accuracy of the calculated values of $a(t - \tau_r(t))$ (see also Ref. 10).

After the calculation of $s(nT_s)$, the new signal sample $a((n + 1)T_s)$ is fetched, interpolation between this sample and the previous one, i.e. $a(nT_s)$, is performed, and the resulting L new signal values are inserted in the circular buffer, the last L values being discarded. That is, only $L - 1$ band-limited interpolations are required for each new microphone output sample. This approach is advantageous when the number of images is larger than the upsampling factor L, as is the case in most applications.

3.3 Solution of non-linear equation

In the moving source problem, the propagation delays $\tau_r(t)$ are defined implicitly as the solutions of the non-linear equation (8). This is in contrast to the fixed source problem where the same equation defines the delays explicitly. Assuming that a particular image will not move considerably during the time interval required for signal transmission from that image to the receiver, it would appear reasonable to approximate $\tau_r(t)$ in eqn (8) with the explicit quantity $\hat{\tau}_r(t)$ in eqn (11). However, even for moderate source velocity, this approximation can introduce important errors in the computed echo arrival times as illustrated by the following example.

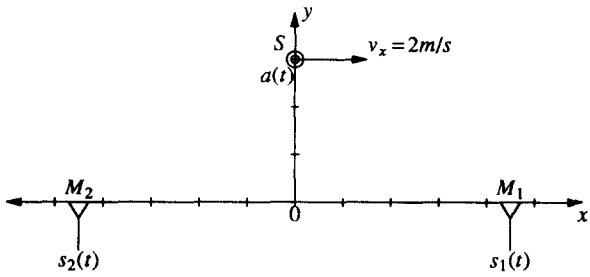

Fig. 1. Geometrical configuration for the example of Section 3.3: a moving source S is monitored with two microphones M_1 and M_2.

Consider the situation depicted in Fig. 1. A source is moving at a constant speed of $v_x = 2$ m/s along a straight-line trajectory described by the position vector $\mathbf{x}(t) = (v_x t, 3 \text{ m}, 0)$. Two microphones located at $\mathbf{x}^{(1)} = (4{\cdot}5 \text{ m}, 0, 0)$ and $\mathbf{x}^{(2)} = (-4{\cdot}5 \text{ m}, 0, 0)$ are used to monitor the source. At $t = 0$, the exact direct path propagation delays (eqn (8)) between the source and each of the microphones are $\tau_0^{(1)} = 1{\cdot}5845 \times 10^{-2}$ s and $\tau_0^{(2)} = 1{\cdot}5692 \times 10^{-2}$ s, respectively. The approximate delays (eqn (11)) are $\hat{\tau}_0^{(1)} = \hat{\tau}_0^{(2)} = 1{\cdot}5768 \times 10^{-2}$ s. Assuming a sampling frequency of 10 kHz ($T_s = 0{\cdot}0001$ s), corresponding errors (shifts) in calculated echo arrival times are $\Delta_0^{(1)} = +0{\cdot}77T_s$ and $\Delta_0^{(2)} = -0{\cdot}77T_s$, respectively.

The above errors in the calculation of the delays are larger in magnitude than the quantization errors introduced by the shifted impulse method at the Nyquist rate, which are upper-bounded by $T_s/2$. The situation is even worse when simulating a rapidly moving source in a large acoustic space since the errors associated with the use of eqn (11) generally increase with the source velocity and with the image distance from the origin. Indeed, even though all the images have the same trace velocity, it takes more time for an echo coming from a distant image to reach the receivers than for an echo coming from a nearby image. During this travel time interval, the distant image will undergo a larger displacement than the nearby image. It is precisely this displacement, which is also proportional to the source velocity, that is not accounted for in the approximation of eqn (11).

The above example indicates that in multi-microphone simulations of moving sources in reverberant space, correct modeling of source motion effects as in eqns (8) and (9) can be considerably more important than the correction of quantization errors with an interpolation technique as in the low-pass impulse method.[5]

In the proposed implementations of eqn (9), Newton's method[11] (also known as the Newton–Raphson method) is used for the solution of the

non-linear equation (8). According to this method, the estimate of $\tau_r(t)$ at the $k + 1$ iteration is obtained as

$$\tau_r^{(k+1)} = \frac{g_r(t - \tau_r^{(k)}) + \tau_r^{(k)} \dot{g}_r(t - \tau_r^{(k)})}{c - \dot{g}_r(t - \tau_r^{(k)})} \tag{12}$$

To evaluate the function $g_r(u)$ eqn (7) and its time derivative $\dot{g}_r(u)$ at any given time u, an algorithmic implementation of the (stationary) image model technique for the particular acoustic space considered is needed. In a typical implementation of this technique, the image position vector $\mathbf{x}_r(u)$ is obtained from the source position vector $\mathbf{x}(u)$ via the application of a series of linear spatial transformations (see Ref. 8). In the present context, the transformations relating $\mathbf{x}_r(u)$ to $\mathbf{x}(u)$ are the same as those relating $\dot{\mathbf{x}}_r(u)$ to $\dot{\mathbf{x}}(u)$, so that $\mathbf{x}_r(u)$ and $\dot{\mathbf{x}}_r(u)$ can be calculated in parallel.

The above approach for the solution of eqn (8) is conceptually simple, requiring only the knowledge of the source position vector $\mathbf{x}(u)$ and its time derivative $\dot{\mathbf{x}}(u)$. As indicated earlier, if $|\dot{\mathbf{x}}_r(u)| < c$ for all u, then eqn (8) has a unique solution $\tau_r(t)$ for any given values of r and t. In addition, if the position vector $\mathbf{x}(u)$ is twice continuously differentiable, so is $g_r(u)$ in eqn (7) and Newton's method is guaranteed to converge, so long as the initial guess is sufficiently close to the desired solution. In this case, the convergence will be quadratic. As explained below, the number of iterations needed for convergence will depend on the proper choice of an initial guess for $\tau_r(t)$, as well as on some external factors related to the source kinematics. In most applications, this number will be relatively small.

A simple initial guess for $\tau_r(t)$ is zero. With this choice, an error analysis of Newton's method[11] yields the following bound on $\varepsilon_r^{(k)} = \tau_r(t) - \tau_r^{(k)}$, the approximation error after k iterations:

$$|\varepsilon_r^{(k)}| \leq D^{-1}(D\tau_{\max})^{2^k} \tag{13}$$

where $\tau_{\max}$ is an upper bound on the travel time of the echoes between the images and the receiver. The constant D is defined in terms of kinematic parameters of the source as follows:

$$D = \frac{\alpha + v^2/d}{2(c - v)} \tag{14}$$

where d represents the minimum distance between the source and the receiver, v is an upper bound on the source velocity, and α is an upper bound on the source acceleration. (A tighter bound can be obtained by using d_r instead of d in eqn (14), where d_r is the minimum distance between the rth image and the receiver). For multi-microphone simula-

TABLE 1
Bound on the Approximation Error $\varepsilon_r^{(k)}$ versus Number of Iterations k

k	*Bound on* $\varepsilon_r^{(k)}$ (*s*)
0	0·25
1	$1{\cdot}22 \times 10^{-4}$
2	$2{\cdot}92 \times 10^{-11}$

tions, the number of iterations k should be such that $|\varepsilon_r^{(k)}| \ll T_s$, where T_s is the sampling interval.

To illustrate the use of eqns (13) and (14), consider again the situation of Fig. 1 and suppose that the reverberation time of the enclosure (not illustrated in the figure) is $T_R = 0{\cdot}25$ s. If only those images within a distance cT_R of the origin are included in the simulation, the choice $\tau_{max} = T_R = 0{\cdot}25$ s is appropriate. For the calculation of D, we set $d = 3$ m, $v = 3$ m/s and $\alpha = 0$. The corresponding values of the bound (eqn (14)) for $k = 0, 1, 2$ are given in Table 1. For an audio signal with sampling frequency 10 kHz ($T_s = 10^{-4}$ s), $k = 2$ iterations are sufficient to achieve the desired level of accuracy.

The number of required iterations can be reduced by using a better initial guess. When $v \ll c$, such a guess is actually given by eqn (11). Note, however, that additional computations are required to evaluate eqn (11) initially. Still another possibility is to use a value of $\tau_r(t)$ calculated at an earlier time, say $\tau_r(t')$ where $t' < t$, as an initial guess in eqn (12).

3.4 Recomputation of image structure

In the fixed source problem, the propagation delays only need to be computed once. In the moving source problem, the propagation delays $\tau_r(t)$ (eqn (8)) are time-varying and (at least in theory) must be recomputed with every new microphone output sample $s(nT_s)$. This poses a serious difficulty when a large number of images are included in the summation of eqn (9), as is usually the case in room simulations. Indeed, accurate calculation of the delays $\tau_r(t)$ for every output sample $s(nT_s)$ adds a considerable computational burden to the procedure, even when only a few iterations of Newton's method are needed to solve eqn (8). Furthermore, as our experience indicates, recomputing the delay structure at intervals of NT_s, where N is a reasonably large integer, and leaving the

delays unchanged between successive evaluations, is not appropriate as important errors can be introduced in the computed echo arrival times. We now present three alterative approaches for efficient and accurate calculation of the delays $\tau_r(t)$ and their time derivatives $\dot{\tau}_r(t)$ as the source moves.

The first approach is referred to as the *time-iterative method.* It is based on the assumption that the delays $\tau_r(t)$ will not change considerably during one sampling interval T_s. In this approach, a single iteration of Newton's method (eqn (12)), with $\tau_r((n-1)T_s)$ as the initial guess, is used in the calculation of $\tau_r(nT_s)$. The derivative of the delay is evaluated simply as $\dot{\tau}_r(nT_s) = [\tau_r(nT_s) - \tau_r((n-1)T_s)]/T_s$. This approach requires M memory spaces, where M is the number of images in the model, to store the delays $\tau_r((n-1)T_s)$ needed in the subsequent calculation of $\tau_r(n\,T_s)$. For slowly moving sources, the time-iterative method converges rapidly and provides very accurate estimates of the propagation delays at all times. However, a single iteration of Newton's method for every image included in the summation of eqn (9) still represents a considerable computational burden when the number of images is large and the function $g_r(t)$ is relatively complex. In such a case, the methods proposed below are recommended.

The second approach, referred to as the *linear interpolation method,* is based on the observation that the delays $\tau_r(t)$ vary smoothly as a function of time. In this approach, the time axis is divided into consecutive frames of duration NT_s samples, where N is an integer referred to as the frame size. The delays are computed exactly at the frame boundaries (i.e. every NT_s samples) using a sufficient number of iterations in Newton's method (eqn (12)). Within a frame, a linear interpolation scheme is used to approximate the delay $\tau_r(t)$. The slope of the interpolation line is used as an approximation to the time derivative of the delay, $\dot{\tau}_r(t)$. Linear interpolation is faster than the single iteration of Newton's method used in the time-iterative method. However, $2M$ memory spaces are required to store the parameters needed for linear interpolation.

An important question is the determination of the maximum value of N that can be used so that the error of linear interpolation at time t, denoted $\varepsilon_r^l(t)$, remains smaller than a preset level. In this case, an error analysis yields the following bound on the interpolation error:

$$|\varepsilon_r^l| \leq \frac{D(NT_s)^2}{2(1 - v/c)^2} \tag{15}$$

where D is defined in eqn (14). As indicated by this bound, the interpolation error usually grows as N^2.

The third approach, referred to as the *quadratic interpolation method*, attempts to extend the number of samples N between successive exact computations of the delays by using quadratic instead of linear interpolation. In this approach, a second-degree polynomial is used to approximate $\tau_r(t)$ over two consecutives frames. The derivative of this polynomial is used to approximate $\dot{\tau}_r(t)$. While this approach is actually effective in extending the frame size N, we note that a quadratic interpolation is computationally more expensive than a linear interpolation. Moreover, $3M$ memory spaces are required to store the parameters needed to carry out the interpolation. This approach is recommended under extreme conditions, i.e. the source is moving rapidly, the calculation of $g_r(t)$ and its derivative is relatively time consuming, and very accurate simulated waveforms are desired.

3.5 Computer implementations

The three methods presented above were implemented in the form of separate computer programs written in the C programming language. A rectangular room model was used for the calculation of the image parameters, i.e. propagation delays and attenuation coefficients. The programs have essentially the same structure, but differ in the way the propagation delays are recomputed, according to the discussion in Section 3.4. The programs are available from the author.

To use any of these programs, it is first necessary to specify the simulated environment, i.e. enclosure dimensions and surface reflectivities. The source position vector, its time derivative, and the receiver position vector are specified in a separate subroutine which is used to calculate the distance function $g_r(t)$ (eqn (7)) and its derivative. Additional parameters related to the band-limited interpolation process must also be supplied, namely the sampling frequency $1/T_s$, the upsampling factor L, the buffer size K and the coefficients of the FIR interpolation filter. In multi-microphone applications, an upsampling factor of $L = 8$ (beyond the Nyquist sampling rate) is usually appropriate. In the present implementation, the FIR filter coefficients are read from an external ASCII file and can be changed as desired.

The programs read the samples of the original source signal from a sequential ASCII file and return the samples of the simulated microphone output signal to a separate ASCII file. In multi-microphone applications, the response of each microphone can be simulated by running the program several times and only changing the receiver position between successive runs. This process can be automated easily. A slightly more efficient approach would be to compute the various microphone

responses in parallel to avoid repeating some of the computations, such as the interpolation of the signal samples $a(nT_s)$. However, the delays $\tau_r(t)$ depend on the receiver position and must be recomputed for each microphone.

While these programs use a rectangular room model in the calculation of the image structure, they can be modified to accommodate more complex geometries. Indeed, the rectangular room model calculation can be replaced by a call to another subroutine that calculates the image position vectors $\mathbf{x}_r(u)$ based on a different image model. For instance, by implementing the algorithm of Lee and Lee[8] in this subroutine, it is actually possible to simulate a moving source in an arbitrary polyhedral enclosure.

4 SIMULATION EXAMPLES

In this section, simulation results are used to illustrate the comparative behavior of the time-iterative, linear interpolation and quadratic interpolation methods. The comparison is made in terms of simulated waveform accuracy and computational requirements. The simulations were performed on a DEC 2100 workstation with 12 mips processing power. The simulation scenario is now described.

We consider a rectangular room with dimensions $L_x = L_y = 10$ m and $L_z = 3$ m, and with wall reflection coefficients $\beta_{x\pm} = 0{\cdot}7$, $\beta_{y\pm} = 0{\cdot}7$, and $\beta_{z\pm} = 0{\cdot}5$. For this room, Sabine's reverberation time is about 0·23 s. The source travels at a constant speed v_y in the y-direction, with its position vector given by $\mathbf{x}(t) = (-4{\cdot}5 \text{ m}, v_y t, 0)$. The receiver is located in the center of the room, i.e. $\mathbf{x}' = (0, 0, 0)$. A zero-mean Gaussian white noise sequence, sampled at 10 kHz (i.e. $T_s = 10^{-4}$ s) and bandpass filtered between 500 Hz and 4·5 kHz is used as the original source signal.

The three programs described in Section 3.5 were used independently to simulate the microphone response. All the images within a distance cT_R from the origin were included in the simulation model. For the band-limited interpolation process (Section 3.2), the upsampling factor was set to $L = 8$ and a 119-coefficient symmetric FIR filter was used. The size of the high-rate interpolation buffer was set to $K = 2300L$. For both the linear and quadratic interpolation methods, $k = 2$ iterations of Newton's method were used for the delay calculation at the frame boundaries.

The simulated waveforms produced by each program were compared to an 'exact' waveform. The latter was obtained by setting the frame size to $N = 1$ in the linear interpolation method. This is equivalent to recomputing the delays $\tau_r(t)$ exactly for every microphone output sample. To

compare the relative accuracy of the three methods, the following signal-to-noise ratio measure was used:

$$\text{SNR} = \sum_n s^2(n) / \sum_n [s(n) - \hat{s}(n)]^2 \tag{16}$$

where $s(n)$ denotes the exact signal and $\hat{s}(n)$ the signal obtained with one of the three methods discussed earlier. Five hundred samples were used for the averaging operation in eqn (16). Note that the SNR is bounded below by -3 dB for completely uncorrelated signal.

Three different values of source speed were considered, namely $v_y = 1$, 2 and 3 m/s. In each case, the time-iterative method produced very accurate waveforms, with SNR on the order of 90 dB. In the current implementation of the program, this figure actually represents an upper limit that cannot be exceeded due to internal representation errors. Moreover, the iterative approach resulted in a factor of 2 saving in computation time when compared to the exact approach (0·61 s/sample for the iterative approach versus 1·22 s/sample for the exact approach).

Figure 2 shows the SNR as a function of the frame size N for the linear and quadratic interpolation methods, when the source speed is set to 1 m/s. It is seen that relatively large values of N can be used without incurring significant distortion in the simulated waveform. For instance, a frame size of $N = 2000$ samples (i.e. exact calculation of the delays every 0·2 s) results in a SNR of 40 dB for the linear interpolation method. As expected, still better results are obtained with the quadratic interpolation method, i.e., SNR = 69 dB for $N = 2000$.

It is important to note that exact calculation of the delays as in eqn (8)

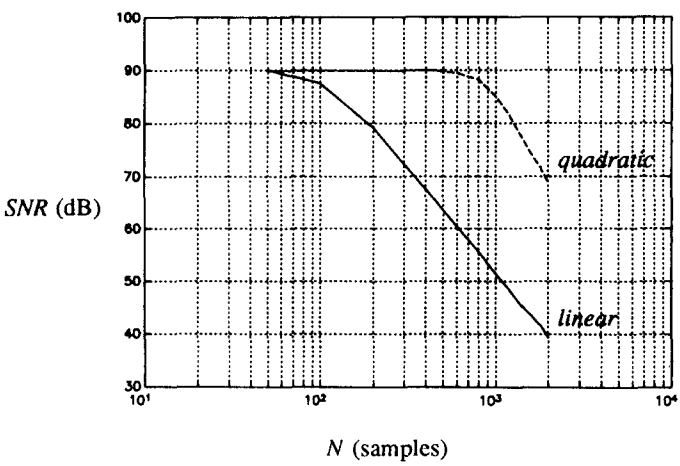

Fig. 2. SNR of simulated microphone output versus frame size N for the linear and quadratic interpolation methods (source speed $v = 1$ m/s).

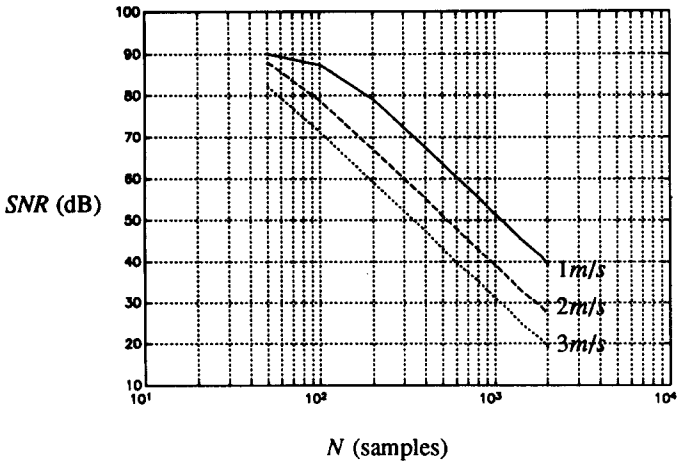

Fig. 3. SNR of simulated microphone output versus frame size N for the linear interpolation method (source speed v = 1, 2 and 3 m/s).

followed by some form of interpolation is essential to achieve such levels of SNR. For example, if the delays are kept fixed during a frame and the approximation of eqn (11) is used to update the delays at the frame boundaries, the SNR values obtained fluctuate between 0 and 5 dB for N between 50 and 1000. These SNR values are inadequate for multi-microphone applications. Still lower values of SNR are obtained with source speeds of 2 and 3 m/s.

Figures 3 and 4 show the effect of increasing the source speed on the accuracy of the simulated waveform for the linear and quadratic inter-

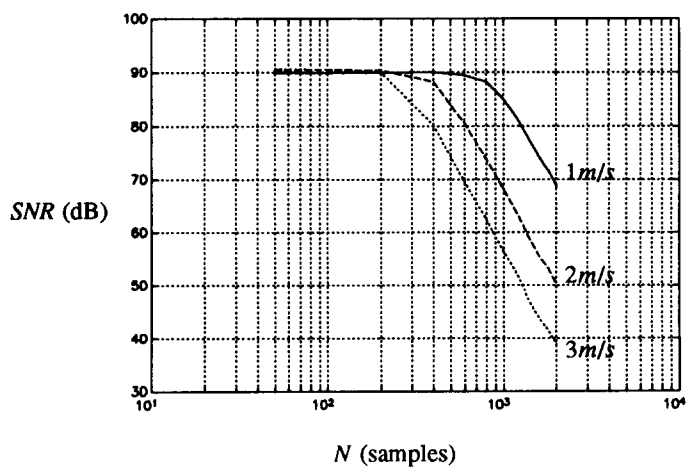

Fig. 4. SNR of simulated microphone output versus frame size N for the quadratic interpolation method (source speed v = 1, 2 and 3 m/s).

polation methods, respectively. As expected, to maintain a constant SNR value, it is necessary to reduce the frame size as the source speed increases.

In this example, the linear and quadratic interpolation methods reduce the computation time by a factor of 5 when compared to the exact calculation (about 0·27 s/sample for the interpolation methods versus 1·22 s/sample for the exact approach). Moreover, a frame size of $N = 100$ is sufficient to achieve this saving. Further increasing N does not reduce the computation time significantly because of the fixed amount of computation necessary to perform linear or quadratic interpolation. However, larger savings are possible in simulations involving complex source trajectories for which the evaluation of $g_r(t)$ is computationally more expensive. In this respect, the potential saving is even larger for non-rectangular geometry where each evaluation of $g_r(t)$ involves a call to a complex algorithm implementing the image model technique for that geometry.

Finally, we note that the fidelity of the simulated microphone response, as measured by the SNR (eqn (16)), depends on the frequency content of the source signal. For instance, a source signal whose energy is concentrated at high frequency will produce lower SNR values than those reported here for a wideband signal. Indeed, the difference between the simulated microphone response $\hat{s}(n)$ and the exact response $s(n)$ is mainly due to small errors in the computed echo arrival times. Obviously, the phase errors corresponding to these arrival time errors are more important at high frequency.

5 SUMMARY AND DISCUSSION

Three new methods have been presented for accurate and efficient simulation of the response of multiple microphones to a moving point source in a reverberant space. The methods, which are based on an extension of the image model technique to a moving source, differ in the way the time-varying image structure is recomputed. They offer different advantages in terms of computational complexity and simulated waveform accuracy, and are particularly appropriate for the study of multi-microphone systems that are sensitive to intermicrophone phase errors. The methods have been implemented in the form of computer programs for a rectangular room model and their comparative performance has been illustrated with simulation results.

The application of the methods is not limited to indoor acoustics. Indeed, by setting some of the reflection coefficients of the boundary

planes to zero, it is actually possible to simulate semi-infinite environments such as those encountered in outdoor and underwater acoustics. Other applications of the methods are also possible. For instance, by using a sequence of impulses as original source signals, the methods can be used to calculate the time-varying impulse response between a moving point source and a fixed receiver. In this respect, the discussion of Section 3.4 on the interpolation methods gives us some indications on the rate at which the impulse response must be recomputed in a given application.

In a recent study on automatic talker-tracking algorithms for microphone arrays,[12] the linear interpolation method was used to simulate the response of multiple microphones to a moving talker in a rectangular conference room. The flexibility of this approach allowed us to test different algorithms under realistic conditions and to investigate the effects of external parameters such as source speed, microphone positions and wall reflection coefficients on the tracking performance.

ACKNOWLEDGMENTS

The author wishes to thank Professor P. Kabal, who made some interesting suggestions concerning the numerical calculation of the propagation delays, and Dr A. Lobo, who wrote an initial version of the computer program for the linear interpolation method.

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada and by a research contract from Bell Northern Research.

REFERENCES

1. Flanagan, J. L., Johnston, J. D., Zahn, R. & Elko, G. W., Computer-steered microphone arrays for sound transduction in large rooms. *J. Acoust. Soc. Amer.*, **78** (1985) 1508–18.
2. Farassopoulos, A., Speech enhancement for hearing aids using adaptive beamformers. *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Glasgow, Scotland, 23–26 May 1989, pp. 1322–5.
3. Goulding, M. & Bird, J. S., Speech enhancement for mobile telephony. *IEEE Trans. Vehic. Technol.*, **39** (1990) 316–26.
4. Allen, J. B. & Berkley, D. A., Image method for efficiently simulating small-room acoustics. *J. Acoust. Soc. Amer.*, **65** (1978) 943–50.
5. Peterson, P. M., Simulating the response of multiple microphones to a single acoustic source in a reverberant room. *J. Acoust. Soc. Amer.*, **80** (1986) 1527–9.

6. Kuttruff, H., *Room Acoustics*, 3rd edn. Elsevier, London, 1991.
7. Borish, J., Extension of the image model to arbitrary polyhedra. *J. Acoust. Soc. Amer.*, **75** (1984) 1827–36.
8. Lee, H. & Lee, B.-H., An efficienct algorithm for the image method technique. *Applied Acoustics*, **24** (1988) 87–115.
9. Morse, P. M. & Ingard, K. U., *Theoretical Acoustics*, McGraw-Hill, New York, 1968.
10. Ramstad, T. A., Digital methods for conservation between arbitrary sampling frequencies, *IEEE Trans. Acoust., Speech and Signal Proc.*, **32** (1984) 577–91.
11. Conte, S. D. & de Boor, C., *Elementary Numerical Analysis: An Algorithmic Approach*, McGraw-Hill, New York, 1972.
12. Lobo, A., Champagne, B. & Kabal, P., On the use of a split-beam array for tracking a moving talker. Paper presented at the 122nd Meeting of the Acoustics Society of American, Houston, Texas, 4–8 Nov 1991.