Synthesis of Room Responses Using the Virtual Source Representation and the Comb-Nested Allpass Filters

Mingsian R. Bai
e-mail: msbai@mail.nctu.edu.tw

Kwuen-Yieng Ou

Department of Mechanical Engineering,
National Chiao-Tung University,
1001 Ta-Hsueh Road,
Hsin-Chu 300,
Taiwan, R.O.C.

An artificial reverberator is proposed in this paper to synthesize room responses. The method employs the virtual source representation and the comb-nested allpass filters to generate the early reflection and late reverberation, respectively, of room responses. The virtual source method is based on a simple representation of sound field with a distribution of discrete simple sources on the boundary. The complex strengths of the virtual sources are then calculated by solving a frequency domain least-square problem. Parameters such as room geometry, size, and wall absorption are naturally incorporated into the synthesis process. The filtering property of human hearing is also exploited in a nonuniform sampling procedure to further simplify the computation. Apart from the early reflection, a comb-allpass filter network is adopted to simulate late reverberations. Optimal parameters of the comb-allpass filter network are obtained using the genetic algorithm (GA). The energy decay curve (EDC) is chosen as the objective function in the GA procedure. Numerical simulations are carried out for a rectangular room and a concert hall model to verify the proposed technique. Subjective listening experiments demonstrate that the present technique is capable of conferring remarkable realism of room responses. [DOI: 10.1115/1.2360951]

1 Introduction

In recent years, reproduction of room effects has received increasing interest of research since it plays an important role in three-dimensional (3D) audio and auralization. Humans' perception of the acoustical settings of environment relies heavily on the cues concealed in the reflections and reverberations associated with the particular room of interest. Without which, spaciousness and externalization would be lost, especially when headsets are used as the means of audio rendering. Room response is indeed one of the most complex types of system dynamics that presents a tremendous challenge to existing software and hardware of audio signal processing. One straightforward approach to create room responses is direct measurement of the impulse response of the room of interest. However, the field measurement method is known to be quite costly and time-consuming to carry out. Alternatively, room responses can be created with the aid of several numerical techniques, ranging from the simple allpass-comb network, the ray tracing algorithm, and the image method to more sophisticated numerical schemes such as the boundary element method (BEM) and the finite element method (FEM). The allpass-comb network is a widely used artificial reverberator for its simplicity [1]. However, it generally requires a tedious procedure of trial and error to adjust the filter parameters before an appropriate design can be reached. The ray-tracing method [2] and image source method [3] are also commonly used for room response simulation. Several refined algorithms along this line have been proposed to improve the precision and efficiency of the calculation [4–7]. However, these methods have some limitations. The image source method becomes complicated when dealing with rooms of arbitrary shapes. The ray-tracing method relies on the high-frequency assumption and is computationally intensive. The FEM or BEM types of methods are based on more rigorous wave acoustics than the forgoing methods. Still, numerical complication and computational loading hinder the methods for practical applications.

In this paper, a new method of room response synthesis is developed in an effort to reach an acceptable compromise between wave acoustics and computational efficiency. The method employs the virtual source representation and the comb-nested allpass filters to generate the early reflection and late reverberation, respectively, of room responses. The early reflection part represents the salient features of room responses with sufficiently high time and frequency domain complexity, whereas the late reverberation part behaves like a highly dense random noise damped with an exponential envelope. In the paper, a technique based on virtual source representation is developed to calculate the early reflection part. The method was inspired by a novel idea suggested by Koopmann et al. who seek to approximate the sound field radiated by an arbitrary source with the aid of the combined fields of a collection of discrete simple sources interior to the surface of the radiator [8]. In our problem, the enclosed sound field in a room is represented by an equivalent combination of the primary source and a finite number of discrete virtual sources distributed on a virtual boundary. The complex strengths of the virtual sources are then calculated using a least-square procedure by solving a matrix inverse problem, with reference to the conditions imposed at the field points on the physical boundary. Consequently, the original problem of finite boundary has been transformed to a free field equivalent so that the numerical complications frequently occurring in FEM/BEM methods can be avoided. This method is also less computationally expensive than the FEM or BEM methods. The geometry and dimensions of the room and frequency-dependent boundary absorptivity can readily be incorporated into the formulation. In addition, the constant-Q filtering property of the human hearing is exploited in the filter synthesis to further simplify the algorithm, without much performance penalty.

Apart from the early reflection part, a comb-allpass filter network is adopted to simulate late reverberations. The filter network is composed of two parallel comb filters cascaded with two-

Contributed by Noise Control and Acoustics Division of ASME for publication in the JOURNAL OF VIBRATION AND ACOUSTICS. Manuscript received October 13, 2004; final manuscript received October 9, 2005. Assoc. Editor: Richard F. Keltie.

layered nested all-pass filters. Optimal parameters of the comb-allpass filter network are obtained using the genetic algorithm (GA). The energy decay curve (EDC) is chosen as the objective function in the GA procedure. As such, the virtual source early reflection module and the GA-optimized comb-allpass network combine to yield natural-sounding room responses in an efficient manner.

The paper begins with a presentation of the virtual source method, followed by the GA optimization of the comb-allpass network. To verify the technique, a simple rectangular room and a realistic concert hall model were employed in the simulation. The reverberators designed by the proposed method are implemented on a personal computer and rendered using a headphone. In addition, subjective tests were also conducted to assess the validity of the presented method. The results and future perspectives are discussed in the conclusions.

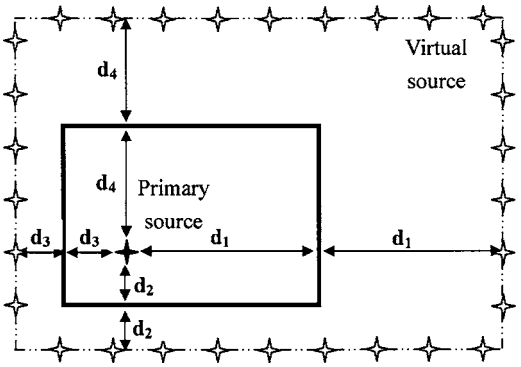


Fig. 1 Vertical view of virtual sources distributed in the neighborhood of the boundary of a rectangular room

2 Modeling of Early Reflection

In this section, the theoretical aspects underlying the present technique will be given with emphasis on the virtual source representation and the specification of boundary conditions.

2.1 Virtual Boundary Source Representation. The method proposed in this paper is motivated by the layer potential theory. It is unnecessary to know both the surface pressure and velocity. The pressure inside or outside of the surface is represented by a simple layer potential [9]

$$p(\mathbf{x}) = \int_S \sigma(\mathbf{x}_0) G(\mathbf{x},\mathbf{x}_0) dS(\mathbf{x}_0) \tag{1}$$

where σ is an unknown strength of the monopole distribution on the boundary. The velocities across the boundary are discontinuous. For $\mathbf{x} \in S$,

$$\frac{\partial p}{\partial n}(\mathbf{x}) = \alpha\sigma(\mathbf{x}) + \int_S \sigma(\mathbf{x}_0) \frac{\partial G}{\partial n}(\mathbf{x},\mathbf{x}_0) dS(\mathbf{x}_0) \tag{2}$$

where

$$\alpha = \begin{cases} 1/2 & \mathbf{x} \in \text{ outside of the boundary} \\ -1/2 & \mathbf{x} \in \text{ inside of the boundary} \\ 0 & \mathbf{x} \text{ right at the boundary} \end{cases}$$

and $G(\mathbf{x},\mathbf{x}_0)$ is the free space Green's function

$$G(\mathbf{x},\mathbf{x}_0) = \frac{A}{r} e^{-jkr} \tag{3}$$

where $(\mathbf{x},\mathbf{x}_0)$ are the field point and the source point, respectively, $\mathbf{r}=\mathbf{x}-\mathbf{x}_0$, $r=|\mathbf{r}|$, A is the source amplitude and $k=\omega/c$ is the wave number, and c is the sound speed. It should also be noted that

$$\frac{\partial G}{\partial n}(\mathbf{x},\mathbf{x}_0) = \frac{\partial G}{\partial r}(\nabla \mathbf{r} \cdot \mathbf{n}) = -\left(jk + \frac{1}{r}\right)\frac{A}{r} e^{-jkr} \cos\theta \tag{4}$$

where $\mathbf{n}$ is the corresponding outward normal and θ is the spanning angle between the outward normal and the vector $\mathbf{r}$. Elaborate numerical schemes such as the FEM or BEM can be used for discretization of the above inherently continuous formulation [10]. Yet, additional insights can be gained by crude numerical integration of Eq. (1),

$$p(\mathbf{x}) = \int_S \sigma(\mathbf{x}_0) G(\mathbf{x},\mathbf{x}_0) dS(\mathbf{x}_0) \approx \sum_{m=1}^{M} \int_{S_m} \sigma(\mathbf{x}_0) G(\mathbf{x},\mathbf{x}_0) dS(\mathbf{x}_0)$$
$$\approx \sum_{m=1}^{M} \sigma(\mathbf{S}_m) G(\mathbf{x},\mathbf{s}_m) \hat{S}_m = \sum_{m=1}^{M} Q_m G(\mathbf{x},\mathbf{S}_m) \tag{5}$$

where $\hat{S}_m$ is the area of the mth element, $\mathbf{s}_m \in \hat{S}_m$, and $Q_m = \sigma(\mathbf{s}_m)\hat{S}_m$ is the strength of the mth monopole. These equations are valid approximation, provided $M \to \infty$, $\hat{S}_m \to 0$. As a consequence, the sound field produced by the actual source is equivalent to a collection of point sources, each with appropriate strength to be determined by additional conditions. All this boils down to the idea that the acoustic field of a complex radiator can be approximated as a superposition of fields generated by a collection of simple sources (monopole, dipole, or combination of both) distributed outside the boundary, which is more tractable numerically than the original integral formulation so that various numerical complications can be avoided [11].

In the present work, a synthesis technique of room responses using a virtual source representation is based on the forgoing idea in which the enclosed sound field is transformed to an equivalent combination of the primary source and a finite number of virtual sources. As motivated by conventional image method, the virtual sources are distributed on a "conformal" surface whose distance to the physical boundary is chosen to be the same as the distance between the primary source and the boundary, as shown in Fig. 1. A field point is chosen correspondingly on the physical boundary for each virtual source, which is the intersection of the boundary with the line connecting the primary source and each virtual source. In contrast to the image method, the listener position does not need to be specified prior to the room response synthesis in the present method.

Since only the impulse response is of interest, we assume for simplicity that the primary source is also a point source of unit strength. According to Eq. (5), the total sound pressure produced by these sources is given by

$$p(\mathbf{x}) = G_p(\mathbf{x},\mathbf{x}_p) + \sum_{m=1}^{M} Q_m G(\mathbf{x},\mathbf{s}_m) \tag{6}$$

where G's are the free space Green's function as defined before, the subscripts p and m signify the primary source and virtual sources, respectively, $\mathbf{x}$ and $\mathbf{s}_m$, $m=1,\ldots,M$ are position vectors, and M is the chosen number of virtual sources that is the same as the number of field points. Note that the number of the virtual sources is chosen empirically. The virtual sources are distributed as uniformly as possible. The rule of thumb is to select virtual sources whose time responses have different time delays. Roughly speaking, the virtual sources fires once at most during each sampling period. Here, Q_m are unknown source strengths to be determined by imposing additional conditions. Thus, direct evaluation of the pressure gradient along the normal direction at the field point $\mathbf{x}$ leads to

$$\frac{\partial p}{\partial n}(\mathbf{x}) = \frac{\partial G_P}{\partial n}(\mathbf{x},\mathbf{x}_P) + \sum_{m=1}^{M} Q_m \frac{\partial p}{\partial n}(\mathbf{x},\mathbf{s}_m) \tag{7}$$

Note that the term $\alpha\sigma(\mathbf{x})$ does not appear in Eq. (7) because the field points never coincide with the source points ($\mathbf{X} \neq \mathbf{s}_m$) and so there is no velocity discontinuity whatsoever as in Eq. (2).

For simplicity, the unknown source strengths Q_m are determined in considering the following *locally reacting* conditions:

$$\frac{\partial p(\mathbf{x})}{\partial n} = -jk\beta(\omega)p(\mathbf{x}) \tag{8}$$

where $\beta(\omega)$ is the *normalized specific acoustic admittance* which is generally dependent on frequencies and boundary conditions. Note that $\beta(\omega)=0$ if the boundary is rigid. Equations (6) and (7) can be rewritten as follows by choosing M field points on the physical boundary of room,

$$\begin{bmatrix} p(\mathbf{x}_1) \\ p(\mathbf{x}_2) \\ \vdots \\ p(\mathbf{x}_M) \end{bmatrix} = \begin{bmatrix} G(\mathbf{x}_1,\mathbf{x}_p) \\ G(\mathbf{x}_2,\mathbf{x}_p) \\ \vdots \\ G(\mathbf{x}_M,\mathbf{x}_p) \end{bmatrix} + \begin{bmatrix} G(\mathbf{x}_1,\mathbf{s}_1) & G(\mathbf{x}_1,\mathbf{s}_2) & \dots & G(\mathbf{x}_1,\mathbf{s}_M) \\ G(\mathbf{x}_2,\mathbf{s}_1) & G(\mathbf{x}_2,\mathbf{s}_2) & \dots & G(\mathbf{x}_2,\mathbf{s}_M) \\ \vdots & \vdots & \ddots & \vdots \\ G(\mathbf{x}_M,\mathbf{s}_1) & G(\mathbf{x}_M,\mathbf{s}_2) & \dots & G(\mathbf{x}_M,\mathbf{s}_M) \end{bmatrix} \begin{bmatrix} Q_1 \\ Q_2 \\ \vdots \\ Q_M \end{bmatrix}$$

and

$$\begin{bmatrix} \frac{\partial p}{\partial n}(\mathbf{x}_1) \\ \frac{\partial p}{\partial n}(\mathbf{x}_2) \\ \vdots \\ \frac{\partial p}{\partial n}(\mathbf{x}_M) \end{bmatrix} = \begin{bmatrix} \frac{\partial G}{\partial n}(\mathbf{x}_1,\mathbf{x}_p) \\ \frac{\partial G}{\partial n}(\mathbf{x}_2,\mathbf{x}_p) \\ \vdots \\ \frac{\partial G}{\partial n}(\mathbf{x}_M,\mathbf{x}_p) \end{bmatrix} + \begin{bmatrix} \frac{\partial G}{\partial n}(\mathbf{x}_1,\mathbf{s}_1) & \frac{\partial G}{\partial n}(\mathbf{x}_1,\mathbf{s}_2) & \dots & \frac{\partial G}{\partial n}(\mathbf{x}_1,\mathbf{s}_M) \\ \frac{\partial G}{\partial n}(\mathbf{x}_2,\mathbf{s}_1) & \frac{\partial G}{\partial n}(\mathbf{x}_2,\mathbf{s}_2) & \dots & \frac{\partial G}{\partial n}(\mathbf{x}_2,\mathbf{s}_M) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial G}{\partial n}(\mathbf{x}_M,\mathbf{s}_1) & \frac{\partial G}{\partial n}(\mathbf{x}_M,\mathbf{s}_2) & \dots & \frac{\partial G}{\partial n}(\mathbf{x}_M,\mathbf{s}_M) \end{bmatrix} \times \begin{bmatrix} Q_1 \\ Q_2 \\ \vdots \\ Q_M \end{bmatrix}$$

or, in matrix form,

$$\mathbf{p} = \mathbf{g}_p + \mathbf{G}\mathbf{q} \tag{9}$$

and

$$\mathbf{p}_n = \mathbf{g}_{np} + \mathbf{G}_n\mathbf{q} \tag{10}$$

By the same token, the boundary condition in Eq. (8) takes the from

$$\mathbf{p}_n = -jk\boldsymbol{\beta}_M(\omega)\mathbf{p} \tag{11}$$

where

$$\boldsymbol{\beta}_M(\omega) = \begin{bmatrix} \beta_1(\omega) & 0 & \dots & 0 \\ 0 & \beta_2(\omega) & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \beta_M(\omega) \end{bmatrix}$$

Finally, substituting Eqs. (9) and (10) into Eq. (11) leads to

$$[\mathbf{G}_n + jk\boldsymbol{\beta}_M(\omega)\mathbf{G}]\mathbf{q}(\omega) = -[\mathbf{g}_{nP} + jk\boldsymbol{\beta}_M(\omega)\mathbf{g}_P] \tag{12}$$

The virtual source vector $\mathbf{q}(\omega)$ can be extracted by solving a matrix inverse problem, using a least-square procedure for each frequency ω.

2.2 Specification of Boundary Conditions. The application of Eq. (12) requires the knowledge of $\beta(\omega)$. A method on the basis of the known room absorption coefficients is proposed to approximate $\beta(\omega)$. In general, the parameter $\beta(\omega)$ can be represented as

$$\beta(\omega) = \frac{\rho_0 c}{z} = \chi + j\eta, \tag{13}$$

where z is the *specific acoustic impedance*, ρ_0 is the air density, c is the speed of sound, χ is the *normalized specific acoustic conductance*, and η is the *normalized specific acoustic susceptance*. The parameter η slightly shifts the modal frequencies (very small in general), whereas χ damps the resonant modes. According to Eq. (13), the specific acoustic impedance z can be written as

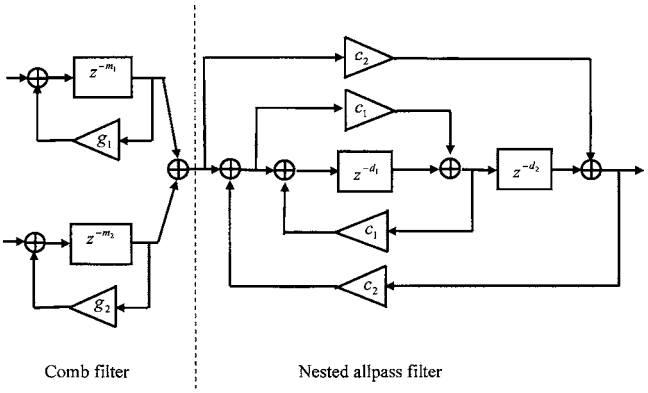
$$z(\omega) = \frac{\rho_0 c}{\chi + j\eta} = r_n(\omega) + jx_n(\omega) \tag{14}$$

where r_n and x_n are specific acoustic resistance and reactance that in turn determines the normal incident sound absorption coefficient of the boundary [12]

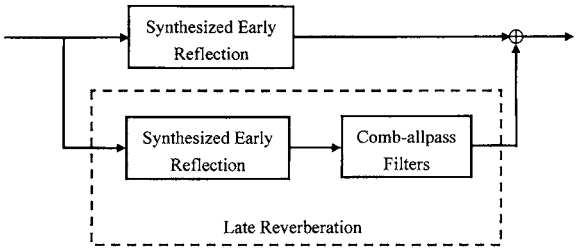
$$\alpha(\omega) = \frac{4\rho_0 c r_n(\omega)}{[r_n(\omega) + \rho_0 c]^2 + x_n(\omega)^2} \tag{15}$$

We assumed that only normal incidence contributes to the early reflection, followed by a diffusive field after a large number of reflections. It is thus plausible to use normal incidence absorption coefficients in the simulation. Further simplification is possible using a crude assumption that $\eta \approx 0$ and $x_n(\omega) \approx 0$. From Eq. (15),

$$z(\omega) \approx r_n(\omega) = \frac{-[2\alpha(\omega)-4]\rho_0 c/\alpha(\omega) \pm \sqrt{[(2\alpha(\omega)-4)\rho_0 c/\alpha(\omega)]^2 - 4(\rho_0 c)^2}}{2} \tag{16}$$



(a)



(b)

Fig. 2 Room response synthesis. (*a*) Block diagrams of the comb-allpass filter network. (*b*) Block diagram of the room response synthesis comprising the early reflection and the late reverberation modules.

Substituting $z(\omega)$ into Eq. (13) gives the value of $\beta(\omega)$. Therefore, $\beta(\omega)$ at each location on the boundary can be derived from the given absorption coefficients at that position. Note that the aforementioned assumption is only expedient and not generally true for absorptive surface. Admitted, this may be one of the factors that contribute to the discrepancies between the measured and the synthesized data. More work needs to be done on how to choose a more appropriate parameter $\beta(\omega)$.

2.3 Nonuniform Spectral Sampling. In the aforementioned procedure, computational loading and synthesis performance is highly dependent on the frequency resolution. A further simplification is possible using a nonuniform sampling scheme that exploits the constant-Q (or proportional band) filtering property of human hearing. The frequency response function is sampled at exponentially spaced points, according to

$$\tilde{\omega}_i = \omega_0 e^{a(2\Pi/P)i},\ i = 0, \ldots, P \tag{17}$$

where ω_0 and P are the lowest frequency and desired number of frequency components. The parameter a must be chosen to cover the desired frequency span $\omega_{\max}$,

$$a = \frac{1}{2\Pi} \ln\left(\frac{\omega_{\max}}{\omega_0}\right) \tag{18}$$

After the weight vectors $q(\tilde{\omega}_i), i=1-p$ are obtained, linear interpolation is called for to resample the weight vectors into the equally spaced frequencies. Then, the frequency response in any position can be calculated using Eq. (6). Finally, the impulse response is obtained from the frequency response utilizing the inverse FFT.

3 Modeling of Late Reverberation

3.1 Structure of the Late Reverberation Filter. Apart from the direct sound field, reverberant sound field is produced by the reflections from the surfaces of the room. After a large number of reflections, the sound in the room may be assumed to have become "diffuse" so that the acoustic energy density is nearly the same throughout the volume of the enclosure, and all directions of propagation are equally probable. In contrast to the early reflections that carry structural information of the room, e.g., the room geometry, absorptivity of the reflecting surfaces, and the position and the directivity of the source, the echo density in the time domain and the modal density in the frequency domain of late reverberation are so high that ordinary modal analysis in wave acoustics becomes impractical. The impulse response of a room is not as structural as the early reflection and may last for seconds with diminishing amplitude, depending only on the room size and absorptivity of the walls. For this reason, instead of detailed modeling, we seek to model the late reverberation by using a simple filter network that allows for producing room responses with similar properties to those of actual rooms. The filter network consists of two parallel comb filters cascaded with two-layered nested allpass filter, as suggested in Gardner's work [13]. Figure 2 illustrates the filter network and the block diagram of room response synthesis. The transfer function of two-layered nested allpass filter is given by

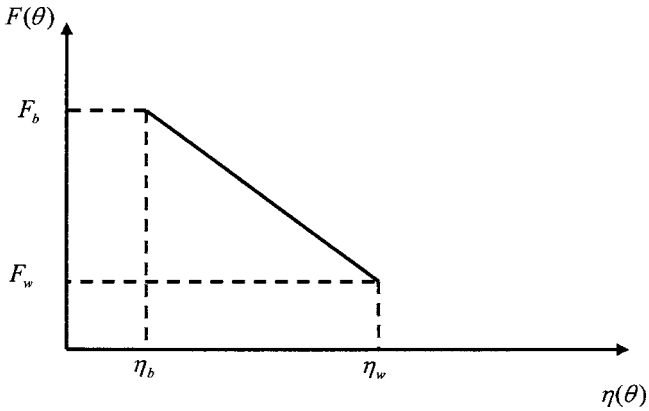


Fig. 3 Relationship between the cost function and the fitness function in GA

$$H_a(z) = \frac{-c_2 + c_1 c_2 z^{-d_1} - c_1 z^{-d_2} + z^{-(d_1+d_2)}}{1 - c_1 z^{-d_1} + c_1 c_2 z^{-d_2} - c_2 z^{-(d_1+d_2)}} \tag{19}$$

where the parameters c_i and d_i $(i=1,2)$ are the gain and delay of the nested allpass filter. There are two noteworthy features of the nested allpass filters. First, the same pattern of response is never repeated in the time domain. Second, the echo density increases with time, which is close to the late reverberation of an actual room response. Next, the transfer function of the comb filter takes the form

$$H_{C_i}(z) = z^{-m_i}/(1 - g_i z^{-m_i}) \tag{20}$$

where the parameters g_i and m_i are the gain and delay of the ith comb filter. It is necessary to determine g_1 of the parallel comb filters and the remaining gain and delay parameters are calculated according to

$$\frac{20 \log 10(g_i)}{m_i T} = \frac{-60}{T_{60}} \quad i = 1,\ 2 \tag{21}$$

where $m_1 : m_2 = 1 : 1.9$ and T is the sampling period.

Despite the simplicity at first glance, it is usually not trivial to choose a set of filter parameters appropriate for a natural reproduction of late reverberation. To minimize the effort of trial and error, a systematic procedure based on the GA is developed in this paper, to be presented in the following section. In the optimization procedure, the reverberation time T_{60} is chosen as the objective

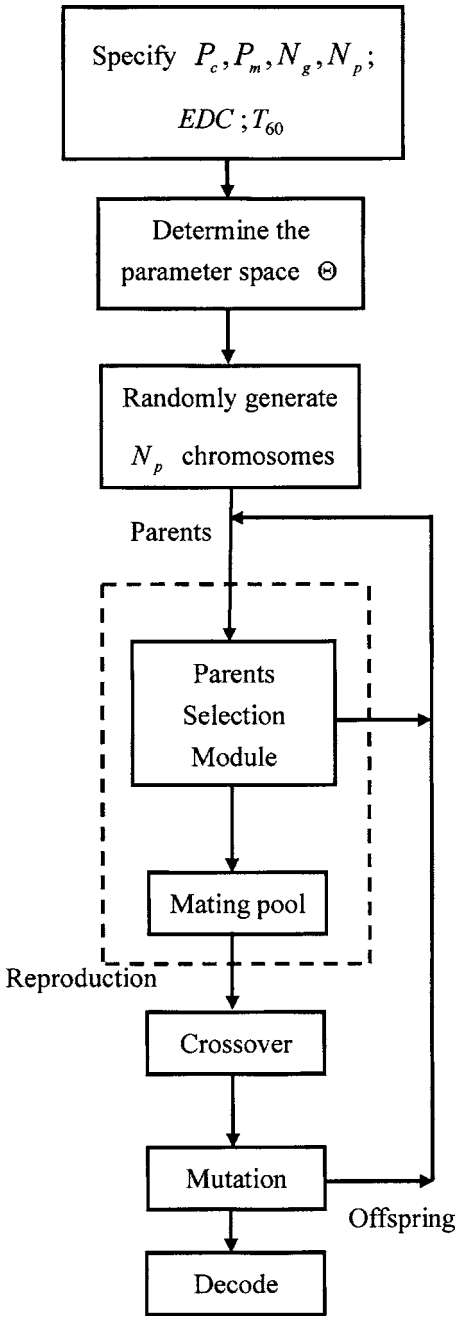


Fig. 4 Flow chart of the GA optimization procedure

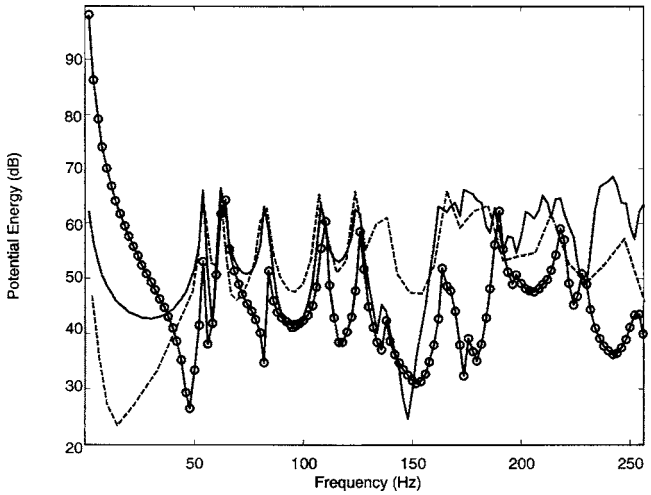


Fig. 5 Comparison of frequency responses calculated using the virtual source representation, BEM; and the analytical solution of the rectangular room. (Solid line: analytical solution; dashed line: virtual source representation; circle line: BEM.)

function that is defined as the time required for the sound pressure level to decay by 60 dB. T_{60} can be estimated on the basis of the energy decay curve (EDC) [14]

$$\mathrm{EDC}(t) = \frac{\int_t^\infty h^2(\tau)d\tau}{\int_0^\infty h^2(\tau)d\tau} \tag{22}$$

where $h(\tau)$ is the impulse response of the room. In the following, a brief review of the GA will be given.

3.2 Genetic Algorithm. The GA is a multi-starting-point optimization algorithm which simulates natural evolution. It is known to not be susceptible to local optima problem, making it well suited for our optimization problem at hand. In the GA procedure, the design variables of Eqs. (19) and (20) are concatenated into the so-called *chromosome*

$$\theta = [g_1 \quad c_1 \quad c_2 \quad d_1 \quad d_2]^T, \ \theta \in \Theta \tag{23}$$

where Θ is the parameter space. The chromosomes encoded as binary numbers are generated randomly to form the population. The objective of the optimization is to find the chromosome with which the EDC of the synthesized room response best approximates the desired one. The cost function $\eta(\theta)$ is defined as

Table 1 The absorption coefficients of the Elmia hall

Surface	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz
Wooden floor (stairs and stage)	0.15	0.08	0.06	0.06	0.06	0.06
Wooden walls and underside of balconies (birch wood)	0.21	0.12	0.09	0.06	0.09	0.13
Ceiling	0.20	0.15	0.10	0.08	0.04	0.02
Reflectors above the stage	0.12	0.10	0.04	0.03	0.03	0.02
Seats in main area	0.45	0.60	0.73	0.80	0.75	0.64
Seats in the rear part	0.50	0.66	0.80	0.88	0.83	0.70
Linoleum floor between the seats (balcony)	0.02	0.03	0.03	0.04	0.06	0.05
Back wall and walls without wood veneer, variable reflectors	0.20	0.12	0.06	0.04	0.07	0.10
Window glass	0.02	0.06	0.03	0.03	0.02	0.02
Plastered concrete (front and underside of speaker rooms)	0.02	0.02	0.03	0.03	0.04	0.06
Ventilation grid on the stage	0.08	0.12	0.15	0.15	0.12	0.08

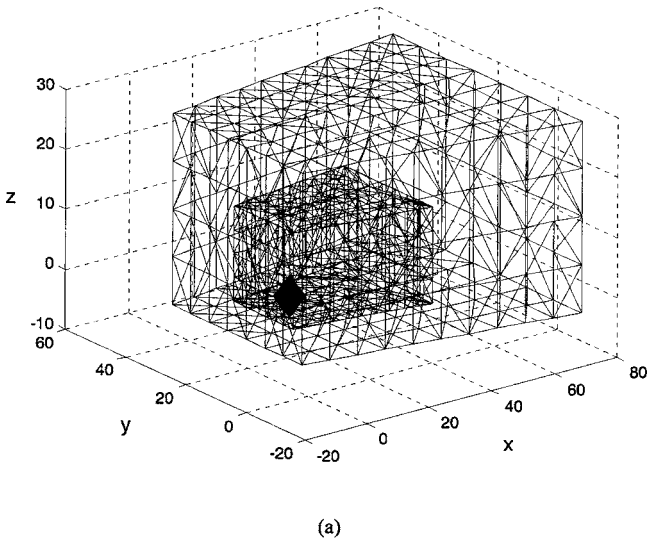


Fig. 6 Virtual sources distribution and the mesh for the Elmia Concert Hall. (*a*) 3D view. The inner and outer meshes correspond to the room boundary and the virtual source surface. (*b*) Vertical view.

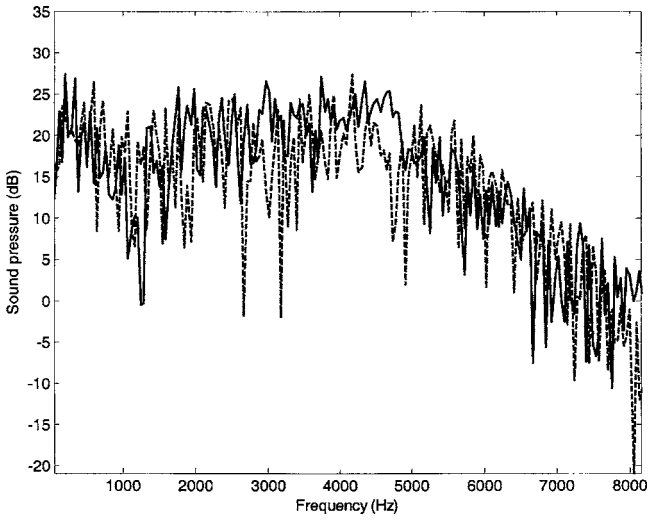


Fig. 7 Comparison of frequency responses calculated using the virtual source method and measured room response of the Elmia Concert Hall. (Solid line: A measured response; dashed line: virtual source method.)

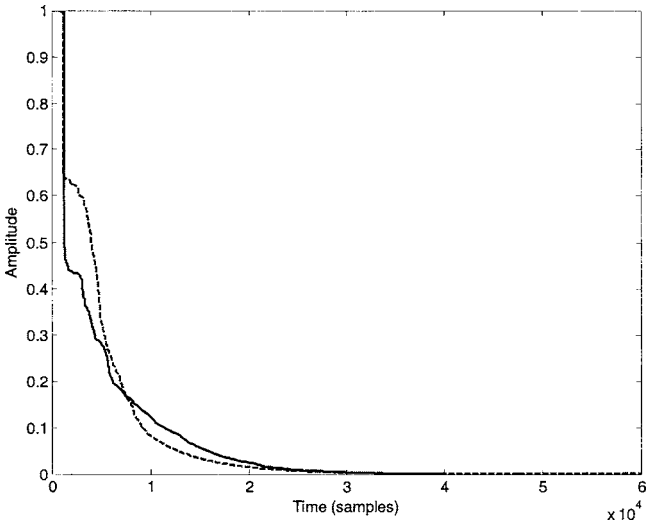


Fig. 8 Comparison of the EDCs resulted from the synthesized and the measured room responses with sample rate 44.1 kHz. (Solid line: measured response; dashed line: virtual source method.)

$$\eta(\theta) = \max_t(|E(t) - \hat{E}(t)|) \tag{24}$$

where $E(t)$ and $\hat{E}(t)$ are the desired and synthesized EDC. The appropriateness of a chromosome is assessed by using the fitness function. According to the principle of the survival of the fittest, a chromosome with higher fitness value is more likely to produce offspring for the next generation. In order to apply the GA to the present optimization problem, the cost function must be related to the fitness function. Referring to Fig. 3, this can be achieved by using the following linear mapping [15,16] also depicted in Fig. 3:

$$F(\theta) = a\,\eta(\theta) + b \tag{25}$$

where

$$a = \frac{F_b - F_w}{\eta_b \eta_w} \tag{26}$$

$$b = F_b - a\,\eta_b \tag{27}$$

The subscripts b and w signify the best and worst values of the cost function, respectively.

In the first GA operation, *reproduction*, the chromosome in current generation will survive depending on the probability

$$p_k = \frac{F_k(\theta)}{\sum_{k=1}^{N_P} F_k(\theta)} \tag{28}$$

where N_p is the population size. The evaluation of the probability distribution function gives the cumulative distribution function (CDF). With reference to the CDF, a random number between 0 and 1 is then generated to decide which chromosome should enter the mating pool. This operation is repeated N_p times until a new

Table 2 Coefficients of the comb and nested all-pass filters

Filter	Comb filter		Nested allpass filter	
Gain	g_1	g_2	c_1	c_2
	0.7078	0.5186	0.4178	0.8733
Delay	m_1	m_2	d_1	d_2
	2001	3802	808	656

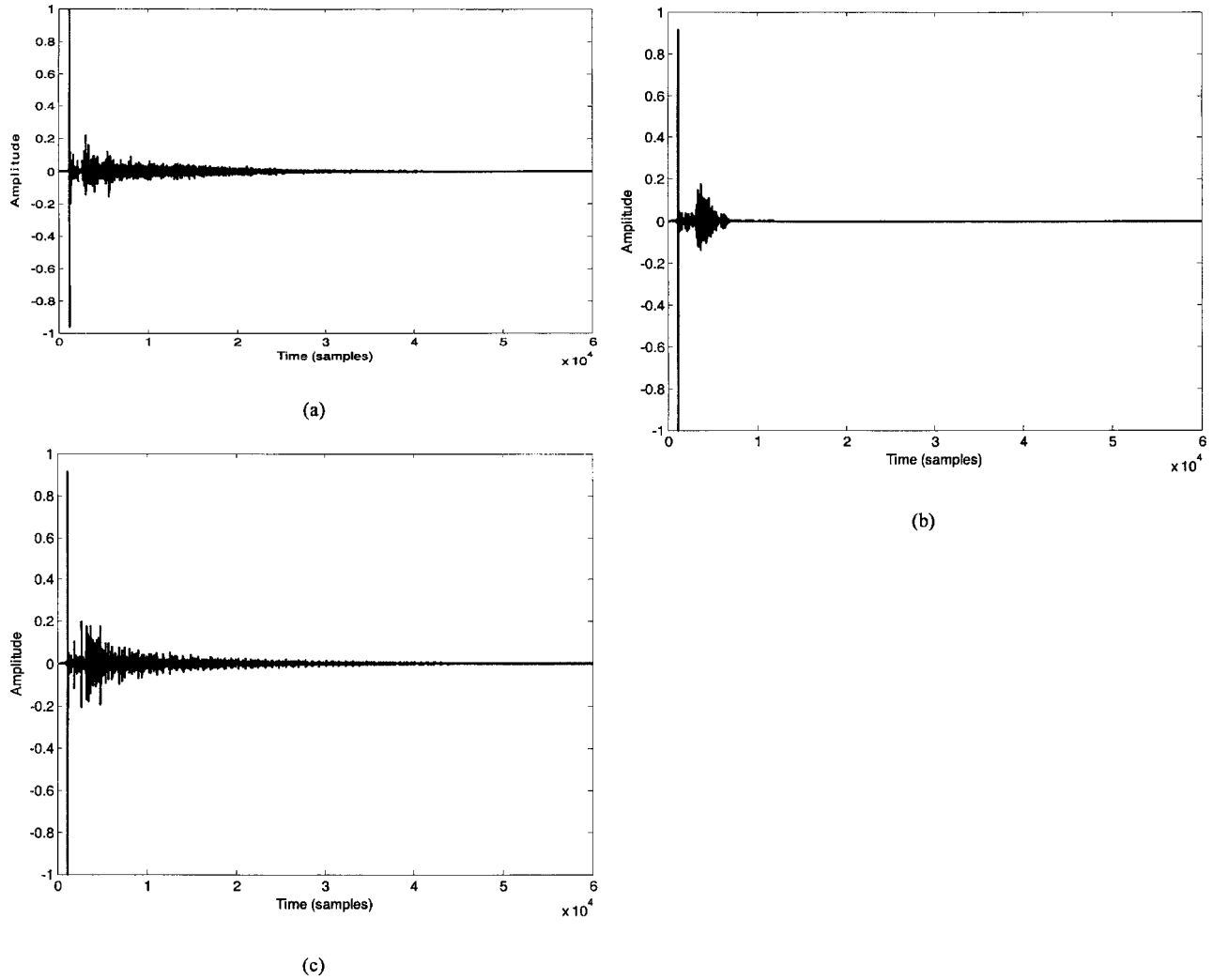


Fig. 9 Impulse responses of the Elmia hall with sample rate 44.1 kHz. (*a*) Measured response. (*b*) Synthesized early reflection. (*c*) Synthesize total response including early reflection plus late reverberation.

population is obtained.

In the second GA operation, *crossover*, two chromosomes are randomly chosen to exchange information in the mating pool. Another random number between 0 and 1 is generated and compared with a prespecified crossover probability (P_c). If this number is less than P_c, the crossover operation will take place.

The third GA operation, *mutation*, is necessary to prevent against premature convergence of the algorithm. This operation is initialized by specifying a mutation probability (P_m). Then, a "masking" string of the length of a chromosome is constructed for each chromosome by randomly generating numbers between 0 and 1. If the number is less than the mutation probability, the corresponding digit in the masking string is set to 1, otherwise, 0. Mutation is then carried out by the "exclusive-OR" operation between the chromosome and the associated masking string. Note, however, according to the elitism policy, two best chromosomes are preserved during the reproduction and passed directly to next generation without crossover and mutation. These GA steps should be repeated until the optimal solution is reached. The overall flow chart of the GA optimization is summarized in Fig. 4.

4 Numerical and Experimental Investigations

4.1 Verification of the Virtual Source Method. Numerical simulations are carried out to verify the aforementioned virtual source method. A small rectangular room model is chosen in the simulation for its known analytical solutions obtained using the modal superposition [11]. Note that the normalized specific acoustic admittance is chosen to be zero ($\beta=0$) because the boundary is rigid. The room dimensions and the locations of primary source and receiver are 3.14 m $\times$ 2.72 m $\times$ 1 m, [1.5 1 0.5] and [2 2 0.8], respectively. The distance between the field points and the virtual sources is the same as the distance between the primary source and boundary. Figure 1 illustrates the distribution of the virtual sources. In addition, the numbers of the field points and the virtual sources are both 202, and $A=1.0$ in Eq. (3). The forgoing nonuniform sampling procedure is employed in the synthesis, with the parameters $\omega_0=2\pi\times 10$ rad/s, $P=64$, $a=0.5161$ in Eq. (17), which leads to $\omega_{\max}=2\pi\times 256$ rad/s. Note that $\omega_{\max}$ is also the sampling rate used for the simulation. In this case, the sampling rate is selected to be 512 Hz. The result calculated using the virtual source method is compared with the analytical solution in Fig. 5. For reference, the simulation result obtained using the BEM is also included. The settings of the BEM are the same as those of the proposed method. In this case, 128 uniformly spaced frequency components are calculated. Using the same Pentium 4 computer, the time required for computation are 37 and 71 s for the virtual source method and the BEM method, respectively.

Comparison of the frequency responses in Fig. 5 reveals that, except for some deviations in level, both methods are capable of capturing the salient features in the low frequency range. It is noted that, with no damping, the analytical and the BEM solutions will go to infinity at resonances. In practical calculations the data

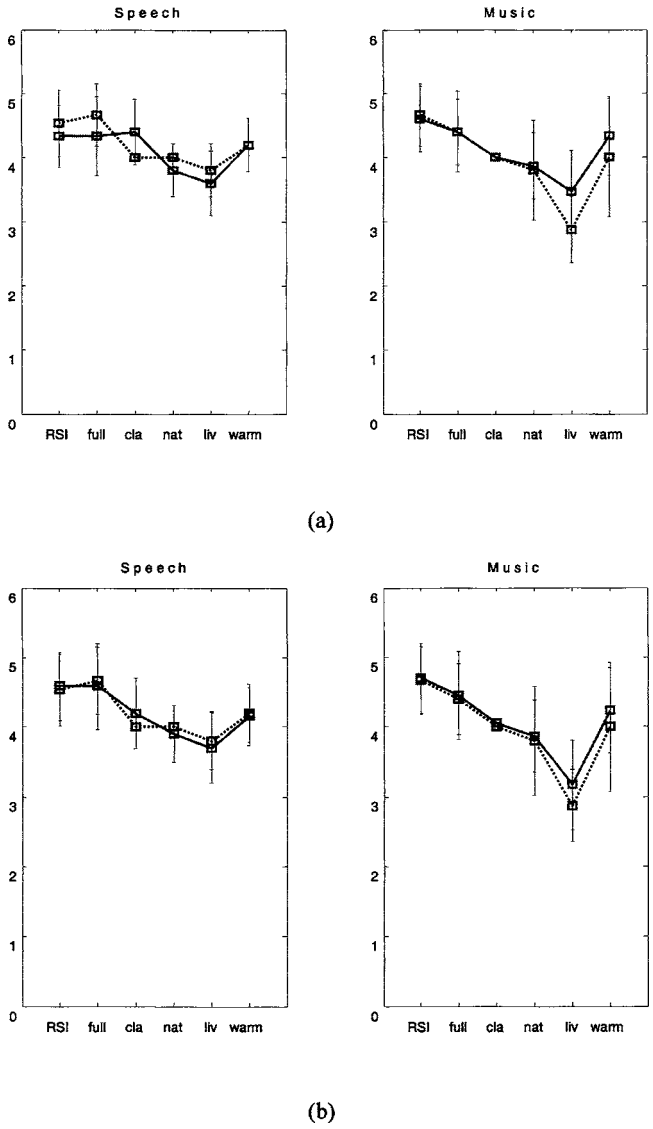


Fig. 10 Comparisons of the audio impression created using the synthesized response and the measured response. (*a*) Synthesized early reflection. (Solid line: synthesized early reflection; dashed line: measured response.) (*b*) Synthesized total response. (Solid line: synthesized total response; dashed line: measured response.) The labels on the horizontal axis are: RSI =room size impression, full=fullness, cla=clarity, nat =naturalness, live=liveness, and warm=warmth.

points will only tend to infinity and be limited by the frequency resolution of the calculations. In order to get a better match, an ad hoc damping (0.01) has been added to the analytical solution. Interestingly, the BEM which is known to be an exact method performed poorly in the prediction of this problem. The main reason is that, in order for all the methods to have an equal basis to compare, nearly identical settings were used in the simulation. Most importantly, the number of field points used in the BEM and the proposed method are the same (202 points). For such limited number of field points, the poor accuracy of the BEM was quite anticipated.

4.2 Optimization. In the following, the Elmia Concert Hall [17] in Sweden is selected to verify the proposed optimized room response synthesizer. Both the virtual source early reflection module and the GA-based late reverberation module are exploited to synthesize the room response. The absorption coefficients of the Elmia hall are shown in Table 1, and the corresponding β can be calculated by Eqs. (13) and (16). The positions of the primary source and receiver are located at [8.5 15 2.1] and [13.8 15 1.45], respectively. The distance between the field points and the virtual sources is set to be equal to the distance between the primary source and the boundary such that the virtual source distribution is "conformal" to the room boundary, as shown in Fig. 6. The number of field points and the virtual sources are both 1412 and the nonuniform sampling procedure is also employed in the synthesis. The parameters $\omega_0=2\pi\times 50$ rad/s, $P=256$, $a=0.8077$ in Eq. (22) such that the $\omega_{\max}=2\pi\times 8000$ rad/s. Such maximum frequency is chosen because the measured data of the Elmia Hall was band-limited to 8 kHz. In addition, the distribution of the source amplitude A was obtained from the website [17]. A low pass filter with cutoff frequency 4 kHz is utilized to simulate the high-frequency absorption of the air. Figure 7 compares the frequency responses of the measured room response with the synthesized early reflection. The result obtained using the virtual source method agrees in general with the measured response. On the other hand, two comb filters cascaded with a two-layered nested allpass filter is used to model the late reverberation. The preceding GA optimization procedure is then employed to optimize the parameters of the comb-allpass filters by matching the EDCs of the synthesized and the measured responses. The parameters used in GA were set to be: population size = 100, number of generations = 100, crossover probability = 0.8, mutation probability = 0.01. Figure 8 compares the EDC of the optimized filter with the desired response, according to the optimal parameters listed in Table 2. In addition, the measured impulse response, synthesized early reflection and synthesized early reflection plus late reverberation are compared in Fig. 9. GA is employed for finding the optimal parameters of the proposed technique, as shown in Fig. 2(*b*). The proportion between the synthesized early reflection and the late reverberation are also optimized in the GA procedure. They can be simply concatenated without any adjustment. The initial peak corresponds to the direct sound. The impulse response in Fig. 9(*a*) lasts for about 40k samples. Figure 9(*b*) indicates that synthesized early reflection is in reasonable agreement with the early portion of the measured response. However, the impulse response lasts for only 10 k samples. The impulse response of the total response shown in Fig. 9(*c*) seems to have captured the general features of the room response. The impulse response lasts for about 40k samples. In order to assess the practicality of the proposed technique, a subjective listening test was undertaken with fifteen participants, using a female speech signal and a symphony signal. Six subjective indices including room size impression (RSI), fullness, clarity, naturalness, liveness, and warmth are employed in the test. Specifically, RSI is the sensation of room size, fullness is the intensity of the reverberation, clarity is intelligibility of the audio signal, naturalness is the fluidity of rendering, liveness is the middle and high frequency reverberation (500 Hz–2 kHz), and warmth is the low frequency reverberation (<500 Hz). Each index is assessed in terms of five discrete levels: 1=very poor, 2=poor, 3=fair, 4=good, and 5 = perfect. The meaning of these subjective indices was fully explained to the participants prior to the listening test. Figure 10(*a*) compares the listening test using the measured response and the synthesized early reflection. The synthesized early reflection performed similarly to the measured response, while notable degradation in terms of spaciousness and fullness were found in the speech case. This suggests that the late reverberation is crucial in subjectively realistic rendering of room responses. Therefore, a listening test was carried out for the synthesized total response (early reflection plus late reverberation). The results are shown in Fig. 10(*b*). It reveals that the total system has indeed achieved excellent performance in comparison to the measured room response.

5 Conclusions

A numerical technique has been developed in the work to synthesize room responses with application in audio reproduction.

The virtual source method is used to calculate the early reflections. It follows the acoustic formalism based on the wave equation and frequency-dependent boundary conditions. Unlike the conventional image method that is often confined to regularly shaped rooms, the present method does not need to prespecify the locations of image points. In addition, late reverberation is synthesized by utilizing two comb filters cascaded with two-layered nested all-pass filters. The parameters of the nested allpass filters as well as the comb filters are optimized using the GA. The proposed technique requires only moderate computation and does not suffer from numerical complications such as the singularity problem. In addition, the constant-Q filtering property of human hearing is exploited in the current synthesis procedure. On the surface, the simulation results showed noticeable discrepancies between the synthesized and the real room responses. The proposed method is capable of producing, not only in the time domain but also in the frequency domain, sufficient response complexity that is close to a practical room. Only salient features of the room response were preserved in the virtual source representation. This has attained the purpose of the present study that seeks to efficiently simulate natural reverberations in the context of human perception, and not the exact reproduction of room responses.

The comb-nested all pass filters have been implemented for simulating the late reverberation. The parameters were optimized by the GA on the basis of the energy decay curve (DEC). The room response synthesizer was completed by combining the early response and the late reverberation modules. As indicated by the subjective listening tests, the present method performed satisfactorily in terms of hearing perception, as compared to the measured response. The proposed technique serves as an efficient and effective means in immersing the listeners into the acoustical environment with remarkable realism.

It is true that the virtual source distribution has significant effects on the accuracy of the approximation. On the other hand, the application of a normal incidence formula to an absorption coefficient that is measured in a diffuse field needs justification. Research on these aspects that could be considered as the limitations of the present work is currently underway.

Acknowledgment

The work was supported by the Nation Science Council in Taiwan, Republic of China, under Project No. NSC 91-2212-E009-032.

References

[1] Jot, J. M., 1992, "An Analysis/Synthesis Approach to Real-Time Artificial Reverberation," *Proc. IEEE, International Conference Acoust., Speech and Signal Processing*, ICASSP-92, San Francisco, **2**, pp. 221–224.
[2] Lee, H., and Lee, B.-H., 1988, "An Efficient Algorithm for the Image Model Technique," Appl. Acoust., **24**, pp. 87–115.
[3] Krokstad, A., Strom, S., and Sorsdal, S., 1983, "Fifteen Years' Experience With Computarized Ray-Tracing," Appl. Acoust., **16**, pp. 291–312.
[4] Drumm, I. A., and Lam, Y. W., 2000, "The Adaptive Beamtracing Algorithm," J. Acoust. Soc. Am., **107**, pp. 1405–1412.
[5] Lewers, T., 1993, "A Combined Beam Tracing and Radiant Exchange Computer Model of Room Acoustics," Appl. Acoust., **38**, pp. 161–178.
[6] Van Maercke, D., and Martin, J., 1993, "Prediction of Echograms and Impulse Responses," Appl. Acoust., **38**, pp. 94–114.
[7] Xiangyang, Z., Kean, C., and Jincai, S., 2002, "A Study of Various Barriers in Enclosed Sound Field by Using the Computer Program-SOFIS," J. Sound Vib., **253**(5), pp. 1115–1124.
[8] Koopmann, G. H., Song, L., and Fahnline, J. B., 1989, "A Method for Computing Acoustic Fields Based on the Principle of Wave Superposition," J. Acoust. Soc. Am., **86**, pp. 2433–2438.
[9] Williams, E. G., 1999, *Fourier Acoustics*, Academic Press.
[10] 1991, *Boundary Element Methods in Acoustics*, R. D. Cliskowski, and C. A. Brebbia, eds., Comp. Mechanics/Elsevier Applied Science, London.
[11] Nelson, P. A., and Elliott, S. J., 1992, *Active Control of Sound*, Academic Press.
[12] Kinsler, L. E., Frey, A. R., Coppens, A. B., and Sanders, J. V., 1982, *Fundamentals of Acoustics*, John Wiley, New York.
[13] Gardner, W. G., 1992, "A Real-Time Multi-Channel Room Simulator," J. Acoust. Soc. Am., **92**(4), p. 2395.
[14] Schroeder, M. R., 1965, "New Method of Measuring Reverberation Time," J. Audio Eng. Soc., **37**, pp. 409–412.
[15] Chen, B. S., and Cheng, Y. M., 1998, "A Structure-Specified H_∞ Optimal Control Design for Practical Applications: A Genetic application," IEEE Trans. Control Syst. Technol., **6**(6), pp. 707–718.
[16] Kozek, T., Roska, T., and Chua, L. O., 1993, "Genetic algorithm for CNN template learning," IEEE Trans. Circuits Syst., I: Fundam. Theory Appl., **40**, pp. 392–402.
[17] Physikalisch-Technische Bundesanstalt, Project 1.401: simulation of room acoustics, http://www.ptb.de/en/org/l/l7/173/roundrob2_l.htm