A Realtime Multichannel Room Simulator

Bill Gardner
Perceptual Computing Group
MIT Media Lab, E15-368C
20 Ames St.
Cambridge, MA 02139
Internet: billg@media.mit.edu
October 30, 1992

Abstract

A room simulator has been developed as part of a project involving virtual acoustic environments. The system is similar to auditorium simulators for home use. The simulated reverberant field is rendered using six loudspeakers evenly spaced around the perimeter of a listening area. Listeners are not constrained to any particular orientation, although best results are obtained near the center of the space. The simulation is driven from a simple description of the desired room and the location of the sound source. The system accepts monophonic input sound and renders the simulated reverberant field in realtime. Early echo generation is based on the source image model, which determines a finite impulse response filter per output channel. Diffuse reverberant field generation is accomplished using infinite impulse response reverberators based on nested and cascaded allpass filters. The system is implemented using Motorola 56001 digital signal processors, one per output channel.

1

Introduction

This paper describes a realtime room acoustics simulator that has been developed as part of the virtual acoustic room project at the MIT Media Lab [Gardner-92]. The basic idea of the virtual acoustic room is to create a room with controllable acoustical properties, using speakers, microphones, and signal processors. The figure below shows the general block diagram of a virtual acoustic room. Sounds created in the physical space are detected by one or more microphones. The microphone signals are then passed through a feedback cancellation system which attenuates speaker originated sounds. The resulting signals represent sounds created within the physical space, and these are passed to the room reverberation rendering system. This system uses a simple room description to synthesize a reverberant field. The resulting array of speaker signals are passed through the feedback cancellation system and broadcast through the speakers.
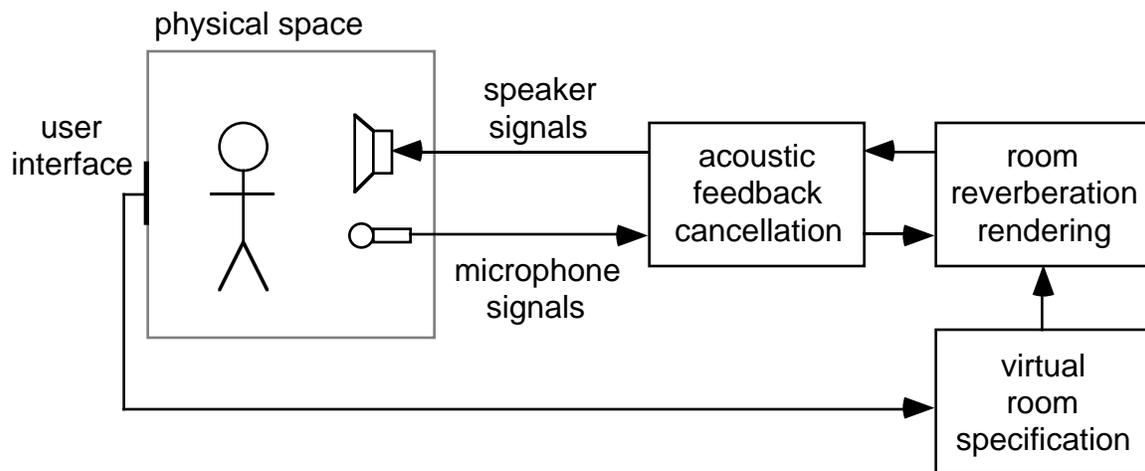
Figure 1 General block diagram of a virtual acoustic room.

Although a complete functional system has yet to be constructed, work is progressing on the major components, notably the feedback cancellation system and the room reverberation system. The feedback cancellation system will most likely use adaptive echo cancellers [Sondhi-92], possibly in conjunction with time-varying reverberation [Griesinger-91]. Note that the latter places some of the feedback cancellation burden on the room reverberation system, thus the boundary between the two systems is fuzzy. An investigation of the combined system would detail the interaction between the two components. This paper, however, will describe a room reverberation rendering system suitable for use in a virtual acoustic room, but treated separately from the feedback cancellation problem. One exception is that the use of echo cancellers constrains the reverberation system to use a small number of microphones and speakers, since the number of echo cancellers required is the product of the number of microphones and speakers.

Problem Statement

The purpose of the room reverberation system is to create speaker outputs that render the input sounds in the context of a specified acoustical space. The figure below shows six speakers surrounding a listener, along with a sound source location and a virtual room perimeter.
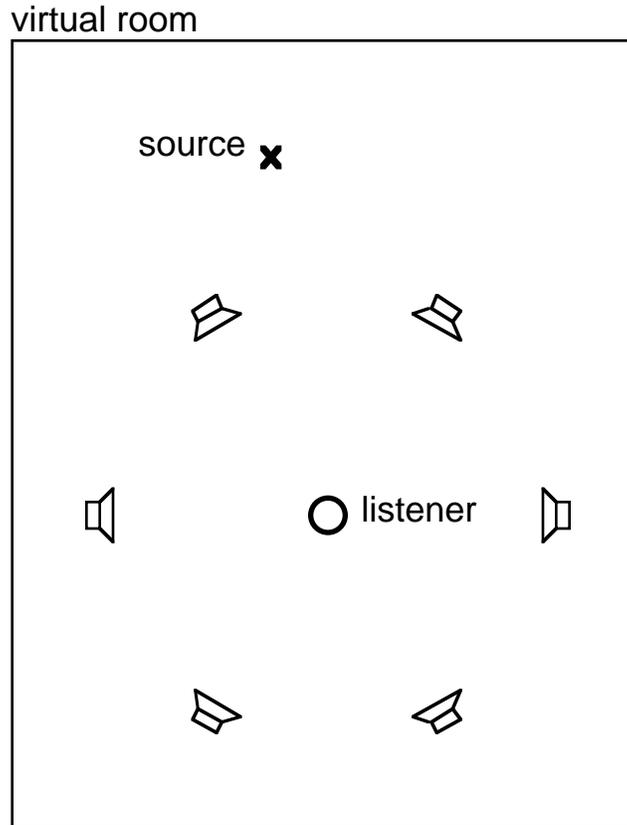

Figure 2. Listener at center of speakers. Sound source is rendered in the context of the virtual acoustic space surrounding the listening space.

The direct sound path from source to listener, the various early reflections off the virtual walls, and the late diffuse reverberation must all be rendered by the speakers surrounding the listener. Note that the listener is not constrained to any particular orientation, although we assume that best listening conditions will occur near the center of the speakers. Thus, rather than using binaural cues to localize the sound source and the various reflections, we will rely on the spatial distribution of the speakers to deliver directional cues.

In the context of an interactive virtual acoustic room, the location of the sound source (microphone signal) would be within the physical space delimited by the speakers. In order to evaluate the reverberation system in a non-interactive way (without microphones), we have chosen to use pre-recorded input sound which is rendered within the virtual space but outside the physical space, as shown in the above figure. With this formulation, the

requirements of the room reverberation system closely resemble those of auditorium simulators for home use [Borish-85] [Griesinger-89]. Two notable differences are that we are assuming the source sound is relatively free from reverberation, and that our listening space is acoustically dead, whereas auditorium simulators for home use must deal with commercial recordings which already contain reverberation, and listening spaces (typically living rooms) which are fairly reverberant as well.

Simulation accuracy is not the main emphasis of this room reverberation system; rather, the goal of this system is to simulate the gross perceptual cues of a variety of typical spaces using a minimum of speakers and processing power, with few constraints on the listener's position. This is in marked contrast to room auralization systems for which simulation accuracy is the primary concern. Such systems generally deliver binaural audio via headphones [Kleiner-91], or via stereo speakers and a head related crosstalk cancellation filter [Schroeder-74]. Other systems have been developed that use many loudspeakers distributed over a hemispherical area in an anechoic chamber [Meyer-65] [Kleiner-81]. Obviously, these systems will do far better at delivering directional cues than a small number of speakers arranged around a listening space.

Six Channel Room Simulator

The reverberation system that has been developed takes a single monophonic input signal and produces six speaker outputs for speakers spaced evenly around a circular area, as shown below.
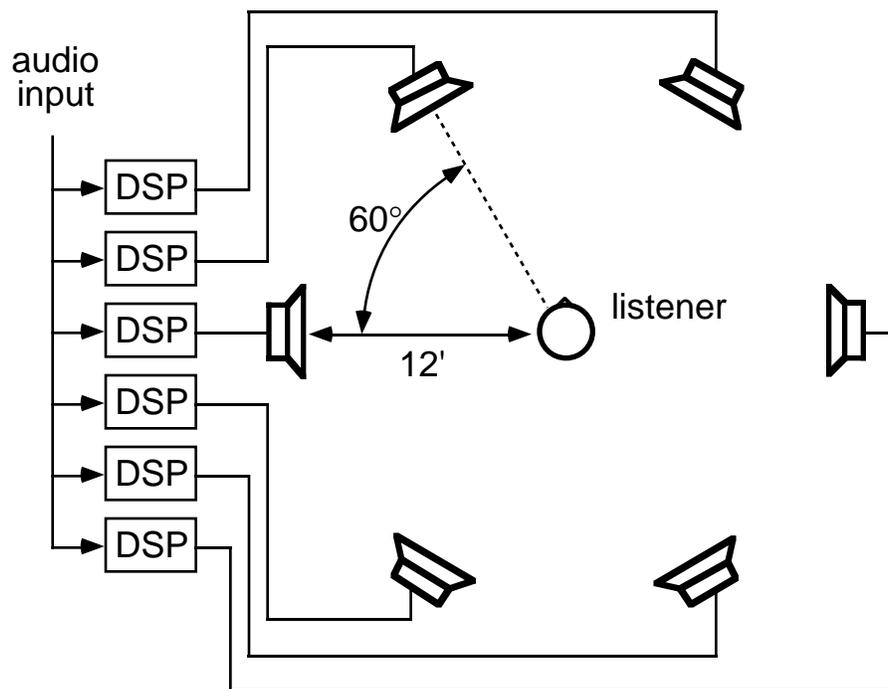
Figure 3 Six channel room simulation system. Speaker height is 5 feet.

The audio input is sent to six digital signal processors (DSPs) which control one speaker each. The DSPs are Audiomedia I cards for the Apple Macintosh computer. The Audiomedia cards are a product of Digidesign Inc., and are based on the Motorola 56000 DSP chip (20 MHz). Two Macintosh II computers are used to host the Audiomedia cards. The computers synchronize virtual room changes via MIDI (Musical Instrument Digital Interface). The system is set up in the Experimental Media Facility (EMF) at the MIT Media Lab. This is not an anechoic environment, but it is quite suitable for non-critical evaluation of this simulator. The room is a very large cube (approximately 60x60x50 feet) and has a broadband reverberation time of 0.61 seconds. This yields a calculated critical distance of roughly 17 feet. The natural acoustics of the EMF were not readily apparent during room simulations, but critical evaluation would require a less echoic setting.

Simulation Overview

The room simulator is driven from a simple description of a room to be simulated. Note that this is merely a convenient way to specify a directional early echo response and a diffuse response, it would be acceptable to use data collected from actual room responses to guide the simulation process. Accurate room simulation requires a detailed geometrical and material description of the room, but for our purposes a much simpler specification is sufficient. Our specification includes the geometry of the perimeter of the room, realized as a polyhedron, the broadband absorption coefficients of the room surfaces, and the locations of the source and listener. Also necessary are the locations of the speakers relative to the listener. This specification allows us to determine the early echo response for the room, which is converted into a finite impulse response (FIR) filter per speaker. The room's volume, surface area and absorption determine the reverberation time of the room via Sabine's equation [Beranek-86]. This determines an infinite impulse response (IIR) filter to simulate the room's diffuse reverberation. The FIR and IIR filters are combined into a single filter specification which is unique per speaker. The filter specifications are then compiled into efficient DSP code, loaded into the DSP cards, and executed. It is possible to quickly choose from different room simulations while listening to source material.
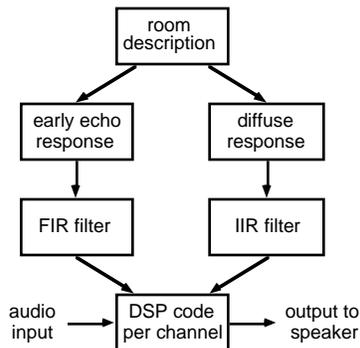

Figure 4 Overview of simulation procedure.

Early Echo Rendering Using the Source Image Method

The source image method models the room as a finite number of polygonal acoustic mirrors. A sound source reflecting off a wall is equivalent to two sources, the original source in front of the wall, and a virtual source (the mirror image of the original source) behind the wall. The source image method can be used to identify all virtual source positions out to a specified maximum distance or maximum number of reflections. The free path propagation from these virtual sources to the listener position then determines the echo response. The figure below shows a corner of a rectangular room containing a source X and a listener O. Some nearby virtual sources are also indicated. From the listener's point of view, listening to the source reflections is equivalent to listening to the free field response of the virtual sources. Finding the virtual sources in arbitrary polyhedral rooms is a complicated, but well understood procedure [Borish-84].
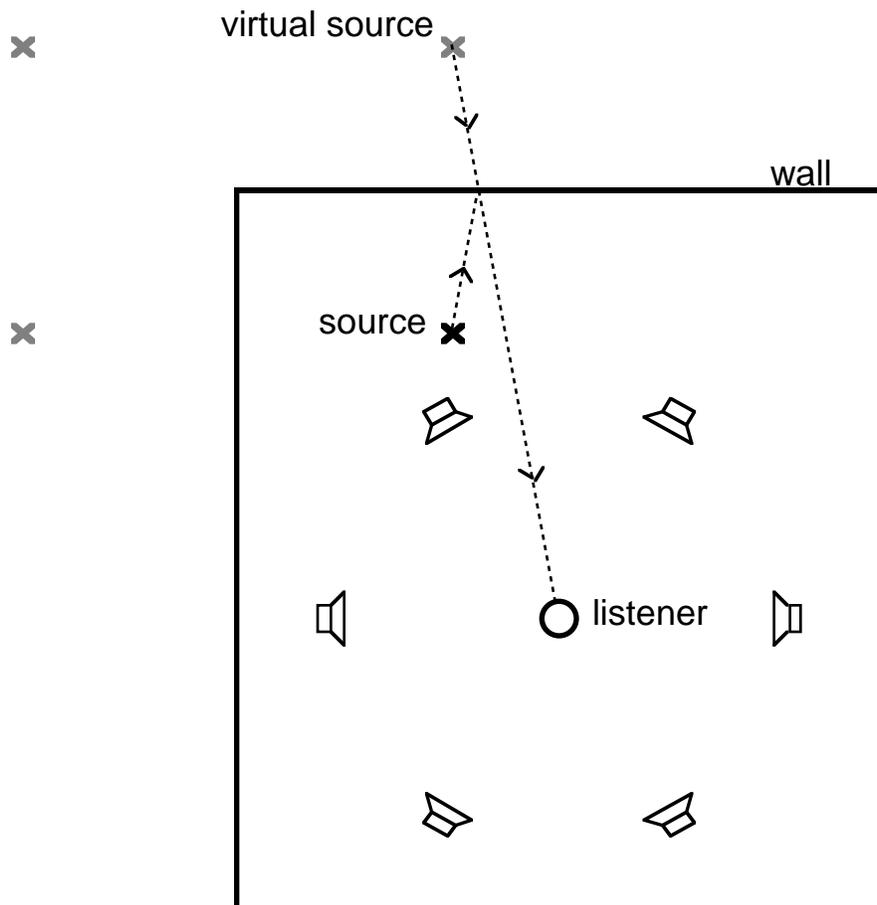

Figure 5 Virtual sources in corner of a rectangular room. The dotted line from the source to the listener represents a reflected sound path which is equivalent to the free field

A program was written that reads a room specification and the source and listener positions and calculates a set of three dimensional virtual source positions using the source image method. All calculations on the sources are then performed in a polar coordinate system with the listener at the origin. All sources not on the horizontal plane (defined as the plane of the speakers and listener) are made horizontal by setting their elevation angles to zero while maintaining their azimuth angle and distance from the listener. During this operation, the amplitudes of the sources are scaled by the cosine of the elevation angle. This way, sources far off the horizontal plane will be ignored, because we have no way to effectively render them, whereas sources close to the horizontal plane are unaffected by this operation. The list of horizontal virtual source positions is then converted to an FIR filter specification for each loudspeaker in the system. The method used relies on intensity panning between adjacent speakers to achieve the desired spatial localization of the virtual sources [Theile-77]. Because the listener is not constrained to any particular orientation, it is unclear how to use phase information to aid in the localization of the virtual sources. The diagram below depicts a virtual source outside the perimeter of the listening space and a listener at the center of the space:
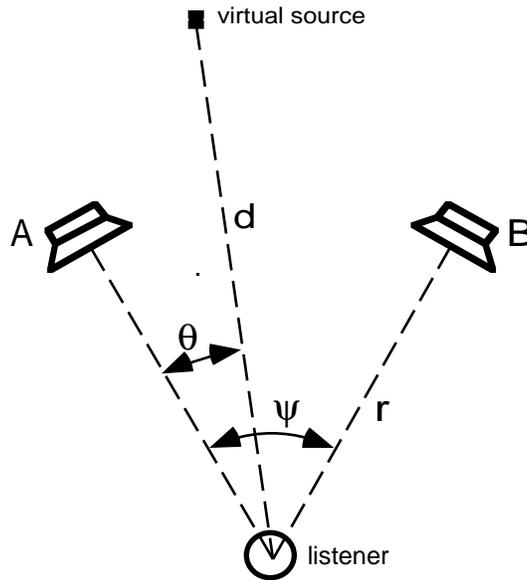

Figure 6 Intensity panning between adjacent speakers.

In the above diagram, the virtual source (with amplitude a) will contribute a filter tap to both the speakers A and B, but to no other speakers. The tap delay lengths depend on the distance from the listener to the virtual source. The tap amplitudes also depend on the distance to the virtual source as well as the angle of the source relative to the speakers:

$$\text{A, B tap delays} \ = \ \frac{d-r}{c} \tag{1}$$

$$\text{A tap amplitude} \ = \ a\frac{r}{d}\cos\left(\frac{\pi\theta}{2\psi}\right) \tag{2}$$

$$\text{B tap amplitude} \ = \ a\frac{r}{d}\sin\left(\frac{\pi\theta}{2\psi}\right) \tag{3}$$

$$a = \cos(\phi)\prod_{j\in S}\Gamma_j \tag{4}$$

where c is the speed of sound, a is the amplitude of the virtual source relative to the direct sound, ø is the elevation angle of the virtual source, S is the set of walls that the sound encounters, and Γ_j is the reflection coefficient of the jth wall. Note that this result applies when the listener, speakers, and virtual source all lie in the same horizontal plane, and the speakers are all equidistant from the listener. A similar result can be derived for the three dimensional case where the speakers are placed on the surface of a sphere with the listener at the center. This would involve panning between more than two speakers at a time.

Note that the rendering of the virtual sources using equal power panning conserves energy. That is, if we were to add speakers, the number of sources per speaker would be reduced, and the overall simulated reverberant energy would stay the same. The attenuation of high elevation sources violates this energy conservation principle somewhat.

Pruning the Early Echo FIR Filter Taps

Typically, the rooms modeled are fairly simple, containing under ten polygonal surfaces. Up to five reflections are calculated, resulting in many hundred virtual sources. The resulting FIR filters contain too many taps to be realized in realtime (40 taps is the maximum), thus pruning the FIR filters is necessary.

Adjacent filter taps within 1 millisecond of each other are merged to form a new tap with the same energy. If the original taps are at times t0 and t1, with amplitudes a0 and a1, the merged tap is created at time t2 with amplitude a2 as follows:

$$t_2 \ = \ \frac{t_0 a_0{}^2 + t_1 a_1{}^2}{a_0{}^2 + a_1{}^2} \tag{5}$$

$$a_2 \ = \ \sqrt{a_0{}^2 + a_1{}^2} \tag{6}$$

Filter taps are then sorted by amplitude, and a prespecified number of the highest amplitude taps are kept. Typically, an all FIR simulation yields 30 tap filters; if diffuse reverb is desired as well, then between 6 and 12 taps are kept per speaker depending on the complexity of the IIR diffuse reverberation algorithm.

The pruning process has the effect of entirely eliminating distant virtual sources, as well as weak taps resulting from intensity panning. Thus, a virtual source that is angularly close to a speaker might be rendered entirely by that speaker after pruning.

Modeling Air Absorption

One improvement to the early echo simulation was to model the frequency dependent absorption of sound by air using a simple one pole lowpass filter. Using the approximations made by [Moorer-79] (at 50% humidity), the following equation was derived:

$$f_c = 2000^{\log_2(d/75)} \tag{7}$$

This equation yields a one pole lowpass cutoff frequency f_c based on the distance of air propagation d in meters. Using this relationship, we can derive a lowpass filter for each FIR filter tap by calculating the echo distance that corresponds to the filter tap. Implementing this strategy is computationally expensive, however. Rather than use a separate lowpass filter for each filter tap, we can use a single lowpass filter for a set of adjacent FIR filter taps by calculating the mean echo distance (weighted by echo energy):

$$\bar{d} = c\frac{\sum_{i \in S} a_i^2 t_i}{\sum_{i \in S} a_i^2} \tag{8}$$

where c is the speed of sound, a_i are the FIR tap amplitudes, t_i are the FIR tap times, and S is the set of adjacent filter taps. Here, for convenience, the calculation is carried out after the virtual sources have been converted to FIR filters.

To minimize computational expense, only one lowpass filter is used per FIR filter, based upon the mean echo distance of the entire FIR filter. Thus, there is a single lowpass filter per output speaker, the exception being that the direct sound FIR taps passed through to the speakers unfiltered. Adding the lowpass filtering to the early echo response improved the simulation considerably, causing the early echo response to sound more natural (i.e. less discrete). Note that a similar filtering mechanism can be used to simulate the frequency dependent nature of reflections, although this has not yet been done.

Diffuse Reverberation Rendering

Moorer determined that an exponentially decaying noise sequence serves as a wonderful sounding impulse response of a diffuse reverberator [Moorer-79]. Rendering this reverberant response requires performing a large convolution. Soon, the price/performance of DSP engines will reach the point where large convolutions can be done in realtime using inexpensive hardware. When this occurs, reverberator implementation will simply be a matter of convolving the input signal with a desired room impulse response, which has either been previously sampled from a real room or synthesized by shaping noise. For the time being, we must be content to implement efficient reverberators for realtime performance. This necessarily implies using infinite impulse response (IIR) filters, such as comb and allpass filters.

Nested Allpass Filters

The trick to designing an efficient, good sounding, diffuse reverberator is to design a linear system whose impulse response resembles a decaying noise sequence. Since white noise has a flat magnitude spectrum but random phase, this suggests the use of allpass filters. Rather than use allpass filters in series as in the Schroeder reverberator, we want to combine them in a way that will lead to an exponential buildup of echoes as occurs in real rooms. One possibility, suggested by [Vercoe-85], is to use nested allpass filters. The idea is to embed an allpass filter into the delay element of another allpass filter. Consider the following flow diagram:
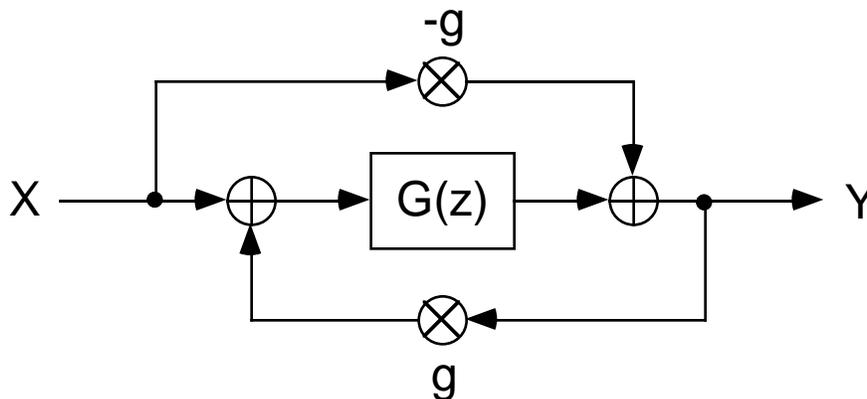

Figure 7 Allpass flow diagram. G(z) must be allpass.

If G(z) is a delay element, this system is a standard allpass filter. The z-transform of this system is given below:

$$H(z) \;=\; \frac{Y(z)}{X(z)} \;=\; \frac{G(z)-g}{1-gG(z)} \tag{9}$$

The magnitude of H(z) is as follows:

10

$$|H(z)| \; = \; \sqrt{\dfrac{|G(z)|^2 - g\bigl(G(z) + G^*(z)\bigr) + g^2}{1 - g\bigl(G(z) + G^*(z)\bigr) + g^2 |G(z)|^2}} \qquad\qquad (10)$$

This equation clearly shows that if the magnitude of G(z) is unity, then the magnitude of H(z) is unity. Thus, H(z) is an allpass system if G(z) is an allpass system. In regards to reverberator design, the advantage to nesting allpass filters can be seen in the time domain. The echoes generated by the inner allpass filters will be recirculated to their inputs via the outer feedback path. Thus, the number of echoes generated in response to an impulse will increase over time rather than remaining constant as with a standard allpass filter.

Because we are using allpass filters, no matter how many are nested or cascaded, the response is still allpass, thus we do not have to worry about stability. It would be possible to nest and cascade comb filters as well, but the response would be highly resonant, and stability would be an issue. It is a mistake to think that because the system is allpass, tonal coloration cannot occur. This is because the short time frequency analysis performed by our ears can detect momentary coloration, and thus allpass systems can sound buzzy, or have a metallic ring, even though they pass all frequencies equally in the long term. A single allpass filter sounds very much like a comb filter; the impulse response is basically a decaying impulse train. When another allpass is inserted into the outer allpass, the impulse response takes on an entirely new character. The number of output echoes increases with time, thus the input "click" is converted into a "pshhh" (or a "bzzzz" with a different choice of delays and gains).

Nested Allpass Implementation

The allpass structure of figure 7 can be implemented easily by attaching operators to a sample delay line as shown below:
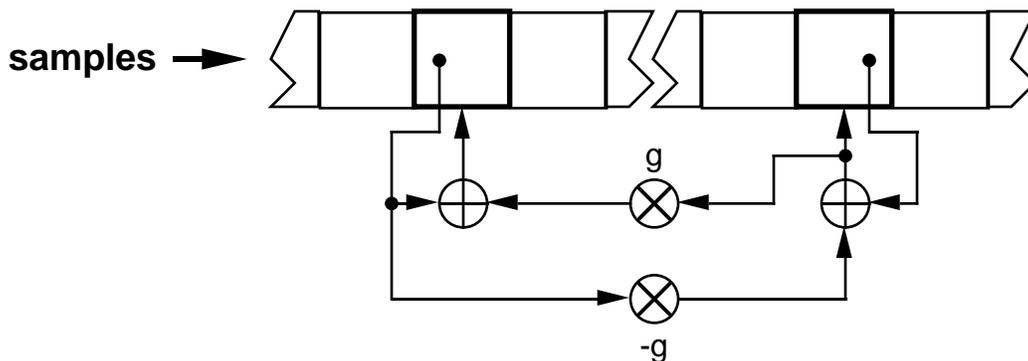

Figure 8 Allpass implementation using a sample delay line.

In the above diagram, the feedforward multiply accumulate through -g occurs before the feedback calculation. After the calculations are complete, the samples in the delay line are shifted one position to the right and processing

continues. Thus, samples entering from the left are allpass filtered and output on the right. In an actual implementation, the samples in memory do not move; instead, the tap locations are shifted to the left, but the effect is the same. This implementation allows us to create arbitrary serial and nested allpass structures with interspersed delay elements by attaching multiple allpass operators to a single delay line. Schematically, this can be represented as follows:
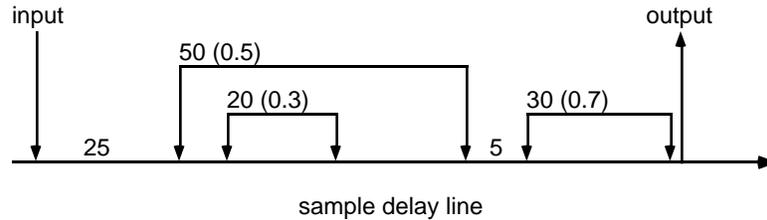

Figure 9 Example of schematic representation of an allpass reverberator.

The above diagram (which is purely instructional) shows the input signal entering a delay line at the left, where it is processed by a double nested allpass cascaded with a single allpass. The element delay lengths are given in milliseconds, and the allpass gains are given in parentheses. Thus, the input signal first passes through 25 milliseconds of delay line, then through a 50 millisecond allpass with a gain of 0.5 that contains a 20 millisecond allpass with a gain of 0.3. Note that because delay elements are commutative, it doesn't matter where the 20 millisecond allpass is located within the 50 millisecond allpass. The output is taken from the delay line after the 30 millisecond allpass. This is called an "output tap". In general, output taps are weighted by a coefficient gain, and multiple weighted output taps may be summed to form a composite output.

Let us consider what happens when the output tap is taken from the interior of an allpass section as shown in the following flow diagram:
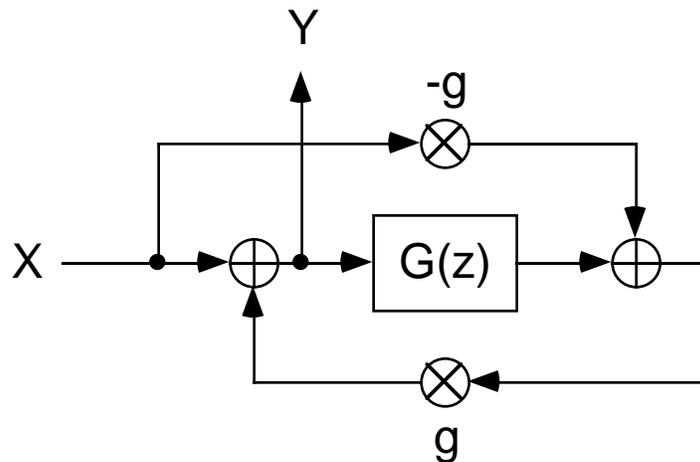

Figure 10 Flow diagram resulting from taking samples from interior of allpass delay line.

12

The z-transform of this system is:

$$H(z) \ = \ \frac{1 - g^2}{1 - gG(z)} \qquad\qquad (11)$$

If G(z) is a delay, then this is a standard comb filter with a constant gain of 1 - g^2, and if G(z) is some other allpass system, H(z) is still a resonant system. If an output tap is taken from the interior of a multiple nested allpass filter, then the resulting system is a cascade of systems of the form in equation 11, and is highly resonant. Experimentation has revealed that these filters sound bad for reverberator design, thus output taps should be taken from locations between cascaded allpasses so that the input/output relationship of each output tap is still allpass. Note, however, that a combination of output taps will not necessarily be allpass because of phase cancellation.

We can use equation 11 to determine how much amplitude headroom we need in the delay lines to prevent overflow within multiple nested allpasses. The magnitude of the system response is:

$$\left| H(z) \right| \ = \ \frac{1 - g^2}{\sqrt{1 - 2g \, \mathrm{Re}\{G(z)\} + g^2 \left| G(z) \right|^2}} \qquad\qquad (12)$$

Since G(z) is allpass, the magnitude of G(z) is unity, and the real part of G(z) can be at most unity, thus the maximum magnitude of H(z) is:

$$\left| H(z) \right|_{\mathrm{max}} \ = \ 1 + g \qquad\qquad (13)$$

Thus, when g is close to unity, the signal within the allpass may be twice the magnitude of the input, and 6 dB of headroom is required. Typically, g is closer to 0.5, requiring only 3 dB of additional headroom per allpass filter.

A General Allpass Reverberator

Despite the attractiveness of these allpass structures for reverberator design, it is difficult to fashion a good sounding reverberator out of simple cascaded and nested allpasses. However, when some of the output of the allpass system is fed back to the input through a moderate delay, wonderful things happen. The harshness, buzziness, and metallic sound of the allpass system is smoothed out, possibly as a result of the increase in echo density caused by the outermost feedback path. This outermost feedback path is essentially a comb filter. A lowpass filter can be inserted into this feedback path to simulate the lowpass effect of air absorption. The general form of this reverberator is given below:
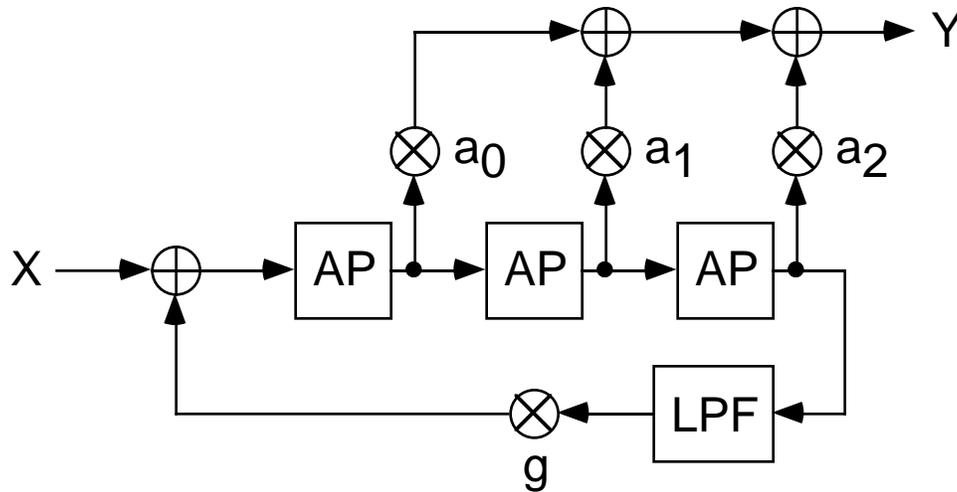
Figure 11 Generalized allpass reverberator with lowpass filtered feedback path and multiple weighted output taps.

The diagram shows a set of cascaded allpass filters with a comb feedback loop containing a lowpass filter. Each of the allpass filters may itself be a cascaded or nested form. Multiple output taps have been taken between allpass sections. This system is no longer allpass, because of the outer comb and lowpass filters, as well as the multiple output taps. However, if the magnitude of the lowpass filter is less than unity for all frequencies, then system stability is guaranteed if g < 1.

As the signal trickles through the cascaded allpasses, each output tap will get a different reverberant response shape. By properly weighting the outputs, it is possible to customize the envelope of the entire reverberator. An adequate lowpass cutoff frequency can be determined by summing the total allpass delay time, converting to a distance by multiplying by the speed of sound, and plugging this "allpass distance" into equation 7, which relates distance to a lowpass filter cutoff frequency. The decay time of the reverberator is controlled by changing g. The decay time can be made extremely long by setting g close to 1. When g is made small, the minimum decay time of the reverberator is limited by the decay time of the allpass sections. However, turning off the outer feedback path (i.e., setting g close to 0) generally causes the response to become gritty and unpleasant.

Obviously, there are a vast number of possible reverberators than can be built with the general structure of figure 11. Unfortunately, it is not obvious how to design such high order filters, especially when the design criteria is simply to sound good. Our ears are very good at detecting patterns in sound. The job of a diffuse reverberator is to elude this pattern recognition mechanism. Thus, perhaps some design criteria can be specified based on avoiding particular recirculating delays which might be easily recognized. Currently, the filter design process has been purely empirical.

14

It should be mentioned that these reverberator structures are not new, although there has been little written about them in the literature. The author's first exposure to them occurred years ago when working in the electronic musical instrument industry. Without a doubt, there are many wonderful sounding commercially available reverberators, all based upon various efficient algorithms. It is unfortunate that the necessities of industrial competition have prevented the open discussion of such algorithms, because they are truly fascinating.
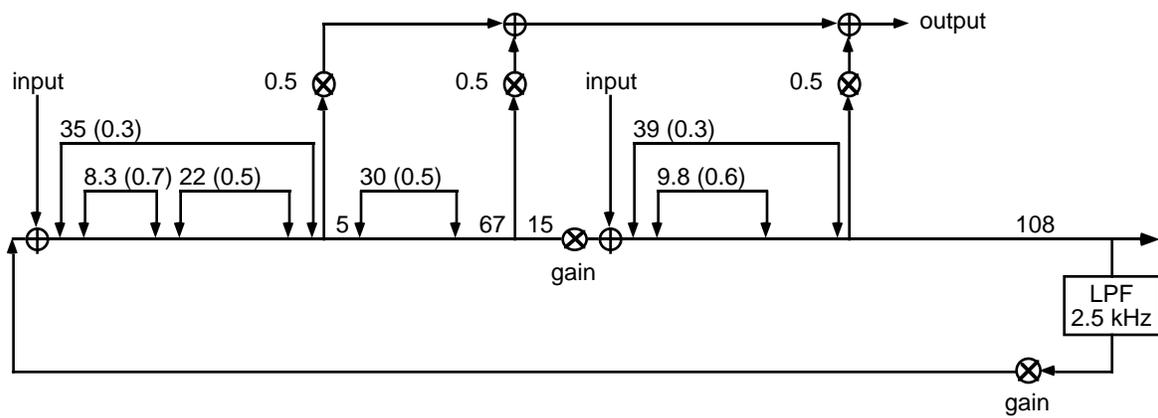
Three Diffuse Reverberators

It was impossible to design a single diffuse reverberator to cover all desired reverberation times. A large room reverberator could not be made arbitrarily small by reducing the feedback gain; similarly, when a small room reverberator was given a large decay time by increasing g, it generally sounded bad. Thus, three different reverberators were designed to cover small, medium, and large rooms. The three reverberators are shown in figure 12. For each reverberator, a mapping was determined between the reverberation time and feedback gain by interpolating between measured data. The table below gives the reverberation time range for each reverberator:

reverberator	RT range (sec)
small	0.38 -> 0.57
medium	0.58 -> 1.29
large	1.30 -> infinite

Small room reverberator:
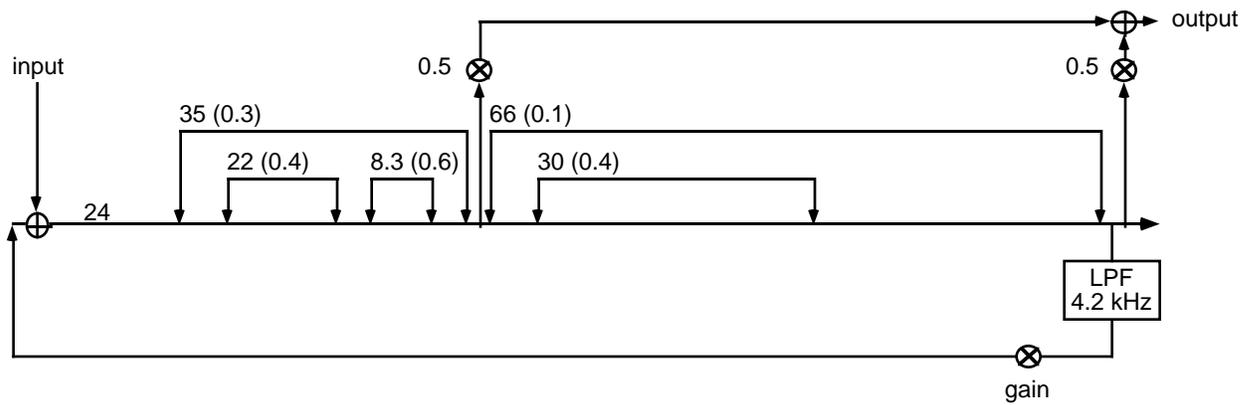
Medium room reverberator:
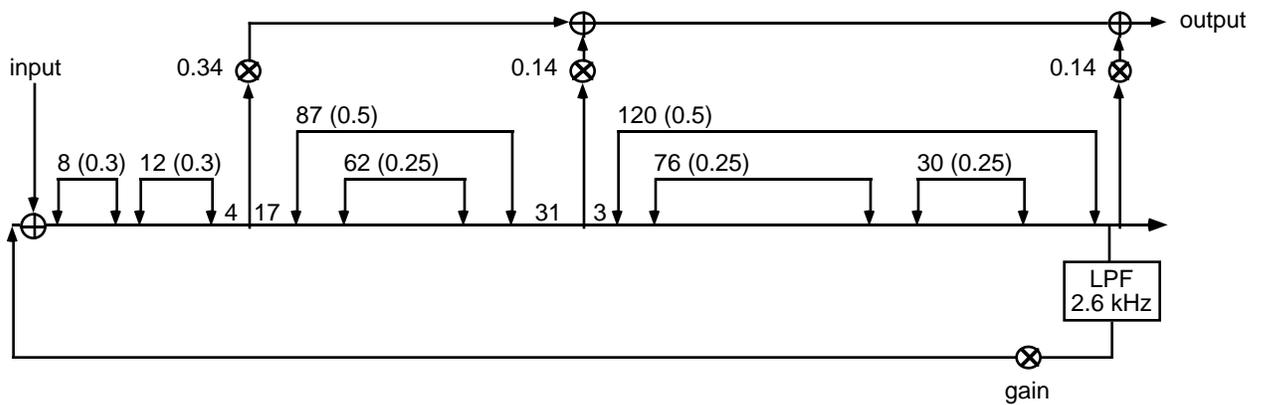

Large room reverberator:

Figure 12 Diffuse reverberators for small, medium, and large rooms. See figure 9 for a description of these schematics.

Creating Spatial Impression

In order to create a diffuse reverberant field that achieves good spatial impression, we need to ensure that the listener receives uncorrelated signals at the two ears. This necessarily requires that the listener receives lateral sound energy, since front-back energy will be correlated at the two ears. Because our system surrounds the listener with speakers, it is sufficient to ensure that the diffuse output of each speaker is uncorrelated with every other speaker. There is a remarkably simple way to do this without redesigning a new reverberator for each channel. By altering slightly all the delay lengths in a reverberator, the new response becomes highly uncorrelated with the original response, even though the gross perceptual qualities remain the same.

For each of the three room reverberators, six variations were created by tweaking the delays slightly. The adjustments to the allpass delays were typically within 2% of the original delay lengths. The variations were auditioned pairwise using headphones to ensure that good spatial impression was achieved between each pair. The final audition was done with the six channel experimental setup using various monophonic music as the source material. The results were excellent, insofar as achieving a surround diffuse reverberant field. The reverberation seemed to come from everywhere, and it was difficult to localize the speakers as being the sound source. Furthermore, the reverberant onset and decays were smooth, so there was no impression of a distinct early echo pattern. The qualities of the three reverberators can be disputed in terms of naturalness and timbre. They are certainly more diffuse sounding than the classic Schroeder reverberator [Schroeder-62], which suffers from a fluttery decay.

Combining Early Echoes with Diffuse Response

The flow diagram given below shows how the early echo FIR filter is combined with the IIR diffuse reverberator for each speaker channel:
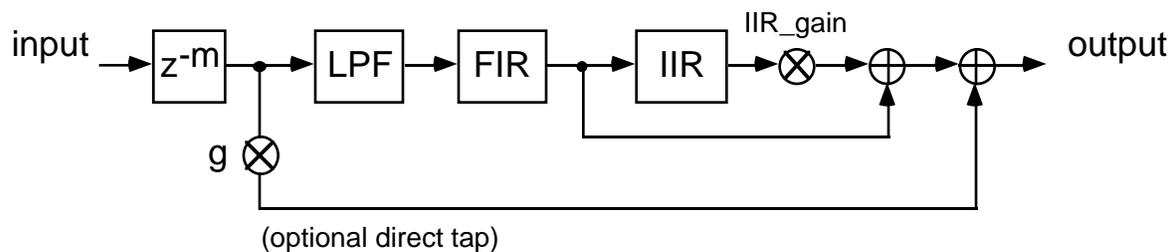

Figure 13 Combining FIR and IIR reverberators.

In the above diagram, LPF represents the early echo lowpass filter, FIR represents the early echo filter, and IIR represents the diffuse reverberator. Note that the diffuse reverberator is driven from the output of the early echo filter, to further increase the echo density. The output is the sum of the early echo response, diffuse response, and optional direct response (which is

unfiltered). The level of the diffuse response is controllable via the IIR_gain multiplier.

The level of the diffuse reverberator needs to be adjusted so that the transition from early echo response to diffuse response is smooth. This can be done by matching the decay slope of the diffuse response with the maximum energy point of the early echo response.
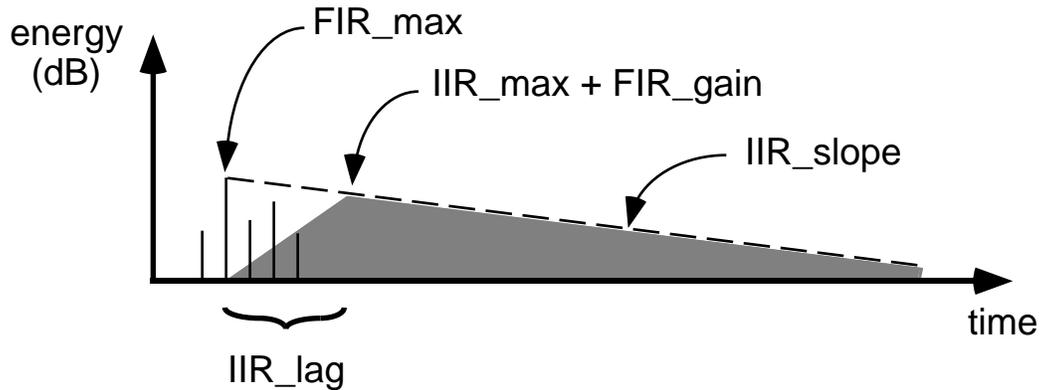

Figure 14 Combining FIR and IIR responses.

The above diagram depicts the FIR early echo response (vertical lines) followed by the IIR diffuse response (gray region). FIR_max is the maximum energy of the FIR response in dB, IIR_max is the maximum energy of the IIR response in dB, which occurs at time IIR_lag seconds after the maximum FIR energy. FIR_gain is the broadband energy gain of the FIR echo response in dB. IIR_slope is simply the reverberant decay slope in dB/sec, and is always negative.

The values IIR_max and IIR_lag are determined a priori for the diffuse reverberator by examining the reverberator response with a nominal reverberation time setting. IIR_slope is determined from the reverberation time of the simulated room which is automatically calculated from the room specification. FIR_max and FIR_gain are determined when the FIR filters are created from the virtual source list, and these values are calculated from the combination of all the FIR filters in ensemble. These values are used to determine IIR_gain as follows:

$$\text{IIR\_gain} \ = \ \text{FIR\_max} + \left(\text{IIR\_slope} \bullet \text{IIR\_lag} \right) - \left(\text{IIR\_max} + \text{FIR\_gain} \right) \quad (14)$$

IIR_gain is the amount we need to raise the diffuse response so that the linear projection of the diffuse response backwards in time will pass through the point of maximum FIR energy. Because we are considering all the FIR responses in ensemble, this determines the IIR_gain setting that matches the overall diffuse level with the combined early echo response from all the speakers.

One remaining issue is that we want the diffuse energy output to be the same from each speaker, corresponding to an omnidirectional diffuse soundfield. However, the diffuse reverberators are driven by the FIR filters which do not have the same energy gains (because the early echo response is direction dependent). Thus, a final adjustment to each channel's IIR_gain is made to ensure the diffuse energy is the same from each channel. The gain adjustments are determined by comparing the energy gain of each channel's FIR filter to the average FIR energy gain. Therefore, this adjustment does not affect the overall diffuse level.

Although this procedure seems complicated, in practice it is straightforward and intuitive. This method of combining the FIR and IIR responses achieves several results, 1) the diffuse reverberator is driven from the early echo response, increasing echo density, 2) the overall diffuse reverberation blends seamlessly with the early echoes, and 3) the diffuse energy output is the same in each channel, even though the early echo energy output differs for each channel.

Summary of Simulation Procedure

The entire procedure for simulating a particular room is as follows:

1) Specify the geometry of the virtual room, and assign absorption coefficients to room surfaces. Specify listener and sound source locations, physical space location within virtual room, and speaker locations.

2) Use source image method to generate virtual source locations. Convert to FIR filters for each speaker. Prune filter taps as necessary.

3) Calculate reverberation time of virtual room, choose proper diffuse reverberator, and determine reverberator feedback gain from empirical relationships.

4) Integrate FIR filters with diffuse reverberators, adjust gains, and compile to final DSP code.

Although some of these steps are currently done by hand, the process is entirely deterministic and could be completely automated.

Results of Simulations

The source material for the simulations is a set of digital classical recordings made by Joe Ierardi using a Kurzweil K2000 digital sampling synthesizer. The recordings are monophonic, extremely dry, and contain a variety of instrumentation and styles. Many different rooms have been simulated with an earlier four channel version of this system, including all of the rooms shown in figure 15.

24' x 32' x 10'   48' x 64' x 15'

RT = 0.7 sec   RT = 1.1 sec

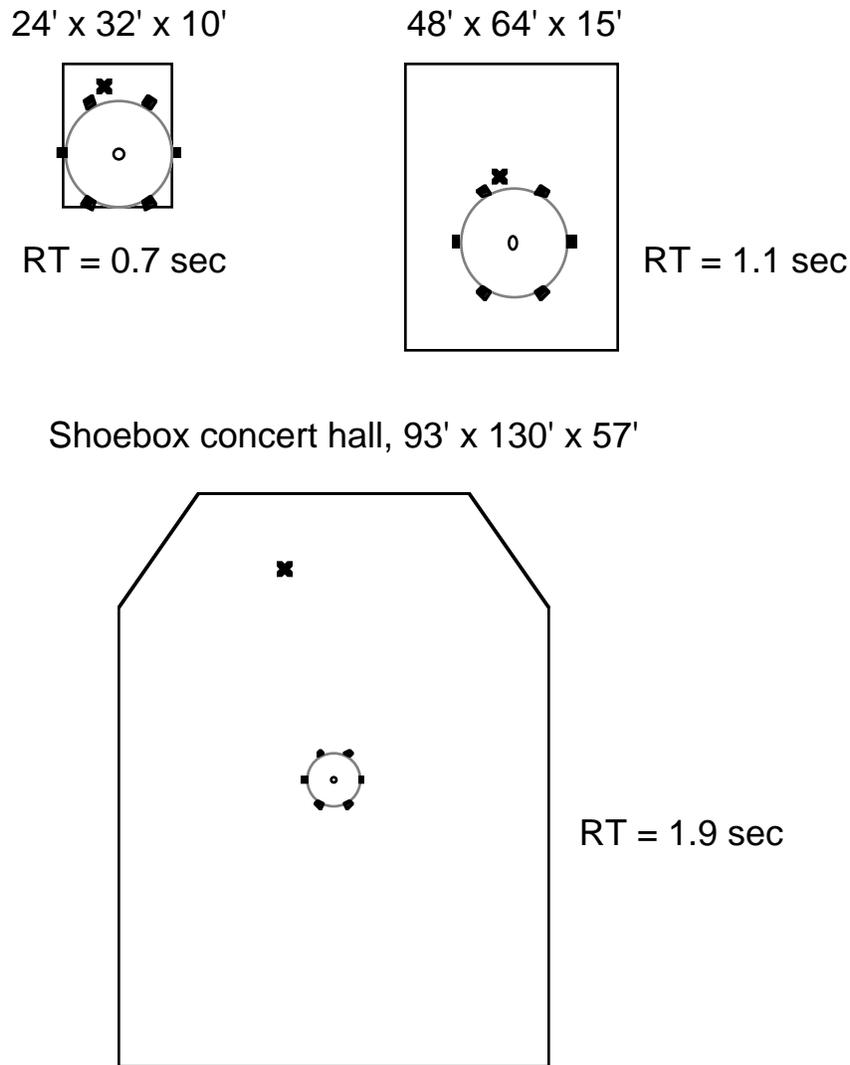Shoebox concert hall, 93' x 130' x 57'

RT = 1.9 sec

Figure 15 Three typical rooms modeled.

As of this writing, only two of these rooms have been completely simulated with the new, six channel system, the small rectangular room and the large shoe box concert hall. The difference between the two simulated rooms is significant. The small room is characterized by the immediate echoes that surround the listener, yet the source is readily apparent at the front of the room. In contrast, the delayed echoes in the large hall cause the energy to slosh around the room, there are distinctive echoes that occur from various directions. These are actually caused by sets of nearby virtual sources, rather than a single isolated reflection. The source distance cues are excellent in both rooms, largely due to the ratio of direct to reverberant energy in conjunction with the room size cues of the early echo pattern. The author and other listeners believe the localization of the source is excellent in the small room, but a few listeners have complained that the sound just seems to be coming from all around. The spatial impression is good in both rooms, the uncorrelated IIR responses cause sound to appear to be coming from all directions, even overhead. Also, it is difficult to localize the speakers in both

simulations, indicating that the combined response is fusing into a single room image. This is only true when the listener is near the center of the space. Note that the simulations done with the six channel system are far superior than those done with the older, four channel system (speakers at corners of square). With the four channel system, it was relatively easy to localize the individual speakers.

Two problems with the simulation are particularly notable. First, the early echo response in the large room suffers from an overly discrete sound, especially in response to an impulsive sound (like a snare drum rim shot). This is clearly due to the simplification in the room model (the use of large planar surfaces) and the lack of sufficient frequency dependent filtering. Second, the diffuse reverberant decay in the large room was somewhat metallic sounding in response to certain input sounds. This sort of problem can only be remedied by tweaking the diffuse algorithms.

Future work

A few areas of future work are indicated below:

• Improve modeling of frequency dependent phenomena. This will necessarily entail adding more filters to the algorithms.

• Compare the reverberation model with actual measured data from a real room. It may be particularly interesting to base the room simulation on real data by fitting the FIR and IIR responses to the directional impulse responses of actual rooms.

• Improve the listening space for critical listening tests. This would involve erecting sound absorbing barriers to create a semi-anechoic listening space.

• Apply psychoacoustical results regarding the perceptual significance of reverberation features to optimize the simulation.

• Continue working on an interactive virtual acoustic room. This entails research into adaptive echo cancellers and time varying reverberation algorithms.

References

[Beranek-86]
Leo L. Beranek, *Acoustics*, American Institute of Physics, New York, NY. (1986).

[Borish-84]
Jeffrey Borish, "Extension of the Image Model to Arbitrary Polyhedra," *J. Acoustical Society of America.* 75 (6) (1984).

[Borish-85]
Jeffrey Borish, "An Auditorium Simulator for Domestic Use," *J Audio Engineering Society*, pp. 330-341 (1985, May).

[Gardner-92]
William G. Gardner, "The Virtual Acoustic Room," master's thesis, Music and Cognition Group, MIT Media Lab, Cambridge, MA. (1992)

[Griesinger-89]
David Griesinger, "Theory and Design of a Digital Audio Signal Processor for Home Use," *J. Audio Engineering Society*, Vol 37, No 1/2, (1989).

[Griesinger-91]
David Griesinger, "Improving Room Acoustics Through Time-Variant Synthetic Reverberation," *J. Audio Engineering Society,* Preprint 3014 (1991).

[Kleiner-81]
Mendel Kleiner, "Speech Intelligibility in Real and Simulated Sound Fields," *Acustica*, Vol. 47, No. 2, (1981).

[Kleiner-91]
Mendel Kleiner, Peter Svensson, Bengt-Inge Dalenback, "Influence of Auditorium Reverberation on the Perceived Quality of Electroacoustic Reverberation Enhancement," *J. Acoustical Society of America.* Preprint 3015 (1991).

[Meyer-65]
von E. Meyer, W. Burgtorf, P. Damaske, "Eine Apparatur Zur Elektroakustischen Nachbildung Von Schallfeldern. Subjektive Horwirkungen Beim Ubergang Koharenz - Inkorarenz," *Acustica*, Vol. 15 (1965).

[Moorer-79]
James A. Moorer, "About This Reverberation Business," *Computer Music Journal*, Vol. 3, No 2 (1979).

[Schroeder-62]
M. R. Schroeder, "Natural Sounding Artificial Reverberation," *J. Audio Engineering Society,* Vol. 10, No 3 (1962).

[Schroeder-74]
M. R. Schroeder, D. Gottlob, and K. F. Siebrasse, "Comparative study of European concert halls: correlation of subjective preference with geometric and acoustic parameters," *J. Acoustical Society of America*, Vol 56, No. 4, (1974).

[Sondhi-92]
Man Mohan Sondhi and Walter Kellerman, "Adaptive Echo Cancellation for Speech Signals," from *Advances in Speech Signal Processing*, edited by Sadaoki Furui and Man Mohan Sondhi. Marcel Dekker, Inc., New York, NY (1992).

[Theile-77]
G. Theile and G. Plenge, "Localization of Lateral Phantom Sources," *J. Audio Engineering Society*, Vol. 25, No. 4, (1977).

[Vercoe-85]
Barry Vercoe and Miller Puckette. "Synthetic Spaces - Artificial Acoustic Ambience from Active Boundary Computation," unpublished NSF proposal (1985). Available from Music and Cognition office at MIT Media Lab.