The Virtual Acoustic Room

by

William Grant Gardner

S.B., Computer Science and Engineering
Massachusetts Institute of Technology,
Cambridge, Massachusetts
1982

SUBMITTED TO THE MEDIA ARTS AND SCIENCES SECTION,
SCHOOL OF ARCHITECTURE AND PLANNING, IN PARTIAL
FULFILLMENT OF THE REQUIREMENTS OF THE DEGREE OF

MASTER OF SCIENCE

AT THE MASSACHUSETTS INSTITUTE OF TECHNOLOGY

SEPTEMBER, 1992

Signature of the Author

Media Arts and Sciences Section
August 10, 1992

Certified by

Barry Lloyd Vercoe, D.M.A.
Professor of Media Arts and Sciences

Accepted by

Stephen A. Benton
Chairperson
Departmental Committee on Graduate Students

The Virtual Acoustic Room

by

William Grant Gardner

Submitted to the Media Arts and Sciences Section, School of Architecture and Planning, on August 10, 1992 in partial fulfillment of the requirements of the degree of Master of Science at the Massachusetts Institute of Technology

Abstract

A room may be used for a wide variety of performances and presentations. Each use places different acoustical requirements on the room. We desire a method of electronically controlling the acoustical properties of a room so that one physical space can accommodate various uses.

A virtual acoustic room is a room equipped with speakers, microphones and signal processors that functions as an interactive room simulator. Sounds created in the room are detected by the microphones, processed to simulate a desired acoustical space, and returned to the room via the speakers. In order to ensure stable operation and enable simulation of arbitrary spaces, acoustic feedback from the speakers to the microphones must be canceled. The resulting system is a combination of acoustic feedback cancellation technology and multichannel room reverberation technology.

This thesis investigates methods applicable to constructing a virtual acoustic room. Acoustic feedback cancellation using static, finite impulse response (FIR) filters is investigated. This technique involves measuring the speaker to microphone response using pseudo-random noise, and creating an FIR cancellation filter from the resulting room response. Multichannel room reverberation rendering is accomplished by using the source image method to determine the early echo response of the virtual room and simulating the diffuse reverberant response using digital reverberators based on nested and cascaded allpass filters. A single channel realtime acoustic feedback cancellation system and a four channel realtime room simulator were constructed.

Thesis Supervisor: Barry Lloyd Vercoe, D.M.A.
Title: Professor of Media Arts and Sciences

This work was supported in part by the Television of Tomorrow Consortium and Pioneer, Incorporated.

Certified by

Judith Brown
Professor of Physics, Wellesley College

Certified by

Bob Chidlaw
Young Chang R&D Institute

Acknowledgements

I would like to thank my advisor, Barry Vercoe, for his unending support for this work. His confidence in me has been most appreciated, especially when I lacked confidence in myself. The same is true of my officemate Dan Ellis, who is a great resource of technical knowledge and a dear friend. I would also thank those people behind the scenes who make it all happen, notably Molly Bancroft and Greg Tucker, who have both put up cheerfully with an endless stream of requests. Thanks go to Bob Chidlaw for introducing me to audio signal processing and thus changing the course of my life, to Malay Kundu for writing the source image software, and to Pioneer for donating speakers and amplifiers. Finally, I thank everyone who has contributed to my success, whether directly or indirectly: my wonderful colleagues here in the Music and Cognition Group, my former colleagues at Kurzweil Music Systems, my roommates, bandmates, friends and family, reader Judy Brown, recommendation writers Dave Mellinger, Don Byrd, and Dennis Picker, and of course, my mother.

Table of Contents

1. Introduction 7
    1.1 Motivation 7
    1.2 Scope of Project 8
    1.3 Organization 10

2. Background 11
    2.1 Room Reverberation 11
    2.2 Reverberation Enhancement Systems 12
    2.3 Reverberation Algorithms 13
    2.4 Room Simulation 15
    2.5 Echo Cancellation 18

3. Acoustic Feedback Cancellation 19
    3.1 Introduction 19
    3.2 Predictive Feedback Cancellation 21
    3.3 Theory 22
    3.4 Energy Decay of Room 25
    3.5 Required Cancellation for a Virtual Acoustic Room 26
    3.6 Cancellation Experiments 29
    3.7 Realtime Cancellation 37
    3.8 Conclusions 39

4. Room Reverberation Modeling 40
    4.1 Introduction 40
    4.2 Early Echo Rendering 42
    4.3 Optimizing the Early Echo FIR Filter 46
    4.4 Results of Early Echo Rendering 46
    4.5 Modeling Air Absorption 47
    4.6 Diffuse Reverberation Rendering 48
    4.7 Nested Allpass Filters 49
    4.8 Nested Allpass Implementation 50
    4.9 A General Allpass Reverberator 53
    4.10 Three Diffuse Reverberators 55
    4.11 Creating Spatial Impression 57
    4.12 Combining Early Echoes with Diffuse Response 57
    4.13 Results of Combined Listening 60
    4.14 The Reverb Compiler 62

5. Conclusions and Future Work 64

References 66

Table of Illustrations

1.1 General block diagram of a virtual acoustic room 9

2.1 Comb filter flow diagram and impulse response 14
2.2 Allpass filter flow diagram and impulse response 14
2.3 Flow diagram of Schroeder reverberator 15
2.4 Virtual sources in a rectangular room 16

3.1 Generalized one channel sound reinforcement system 19
3.2 Predictive feedback cancellation 21
3.3 Impulse response of typical room 23
3.4 Predictive feedback cancellation 23
3.5 Amplitude envelope of idealized room response 25
3.6 Block diagram of experimental cancellation setup 30
3.7 Measured impulse response and energy contour of office 33
3.8 Cancellation results for noise and speech signals 34
3.9 Calculated cancellation results and actual results 36
3.10 Realtime cancellation block diagram 38

4.1 Four channel experimental audio system 41
4.2 Intensity panning between adjacent speakers 42
4.3 Direction of phantom source versus interchannel level-difference for the lateral loudspeakers of a quadraphonic arrangement 44
4.4 Perceived versus desired sound direction with noise signals, using 6 speakers arranged at 60 degree increments 45
4.5 Rectangular virtual space and direct source location 45
4.6 Allpass flow diagram 49
4.7 Allpass implementation using a sample delay line 51
4.8 Example of schematic representation of an allpass reverberator 51
4.9 Flow diagram resulting from taking samples from interior of allpass delay line 52
4.10 Generalized allpass reverberator 54
4.11 Diffuse reverberators for small, medium, and large rooms 56
4.12 Combining FIR and IIR reverberators 58
4.13 Combining FIR and IIR responses 58

1. Introduction

1.1 Motivation

The motivation for this project stems from the importance of room reverberation as it relates to the listening experience. Typically, when we listen to a sound in a room, most of the sound we hear is reflected sound. By containing and reflecting the sound energy, a room increases the sound level and makes acoustic performance to an audience possible. We can think of the room as a signal processor inserted between the sound source and the listener, which affects the level, envelope, timbre, and spatial impression of the original sound, rendering it within the context of an acoustic space. We speak of a room as having "good acoustics" if the effect of the room contributes positively to the listening experience. This is entirely dependent on the type of sound generated and the particular use of the room at the time. In many cases, the effect of the room is not at all subtle; the room can ruin a performance or greatly improve it. Important examples where the room's acoustical properties have a significant role include lectures, media presentation (e.g. cinema, television), theatre, and musical performance ranging from solo recitals to jazz and rock bands to symphony orchestras. All of these uses place different acoustical requirements on the room.

When a performance space is designed, its acoustical properties are targeted for a particular use, or a small range of uses. Consequently, one must seek out a proper acoustical space for a particular performance. In addition, the vast majority of architectural spaces are designed with no regard to acoustical properties. Thus, in living and working spaces we are stuck with what we get acoustically.

A wonderful solution to these problems would be a room with electronically controllable acoustics. Performance spaces could be tailored to each particular use, thus a single performance space could accommodate a variety of functions. Perhaps more significantly, living and working spaces could be acoustically customized to the desires of the occupants. This is not a frivolous idea. For many people, the ability to adjust the acoustic parameters of their personal space would be welcome indeed. Consider, for example, a musician rehearsing a piece in an apartment, but hearing the acoustics of a concert hall.

I introduce the concept of a virtual acoustic room, a room designed to have electronically controllable acoustical properties, and yet function like an ordinary room. I will make the distinction between the physical space of the room and the synthesized virtual acoustic space that surrounds the physical space. The implementation of the virtual acoustic room will necessarily include microphones to detect sound created in the room, signal processors to simulate a desired acoustical space, and loudspeakers to return the processed sound to the physical space.

1.2 Scope of Project

This thesis investigates methods applicable to constructing a functional virtual acoustic room. I will consider a system that uses speakers and microphones located at the perimeter of the physical space. To make the task easier, I will assume that the physical space in which the system is installed is acoustically neutral, so the naturally occurring acoustical properties do not overwhelm the synthetic acoustics. Note that causality limits the size of the virtual acoustic space to be no smaller than the physical space, therefore it is impossible to simulate small room acoustics with a large physical space. Because of this, and the desire to simulate a variety of different sized rooms, I will assume the physical space is relatively small. In addition, I will ignore the usual performance paradigm of a performer on stage before an audience; in this instance, anyone in the virtual acoustic room is both a performer and a listener.

The premise of this thesis is that the implementation of the virtual acoustic room requires the solution to two independent problems: cancellation of the acoustic feedback between the speakers and microphones in the physical space, and the rendering of the reverberant field of the simulated acoustic space.

Acoustic feedback from the speakers to the microphones must be prevented. Unchecked, it will result in either unstable operation or coloration of the resulting reverberant field, or it will force operation at insufficient gain levels to be useful. Furthermore, acoustic feedback prevents the system from simulating arbitrary spaces because the system hears and reprocesses its own output. For the acoustic feedback cancellation problem, I will investigate a solution based upon predictive cancellation. Each speaker microphone pair is modeled as a linear, time-invariant system whose system response can be measured. This enables us to predict what sound will arrive at a given microphone originating from a given speaker. Thus, we can cancel any speaker originated sound arriving at the microphones.

The reverberation rendering problem is somewhat easier, and can be solved adequately by modeling the early reflection portion of the reverberant field separately from the later diffuse reverberation.

The figure below shows the general block diagram of a virtual acoustic room. Sounds created in the physical space are detected by an array of microphones, these are passed through a feedback cancellation system which removes speaker originated sounds. The resulting microphone signals contain sounds created within the physical space, and these are passed to the room reverberation rendering system. The resulting array of speaker signals are passed through the feedback cancellation system and broadcast through the speakers. The simulated room is selected by a suitable user-interface.

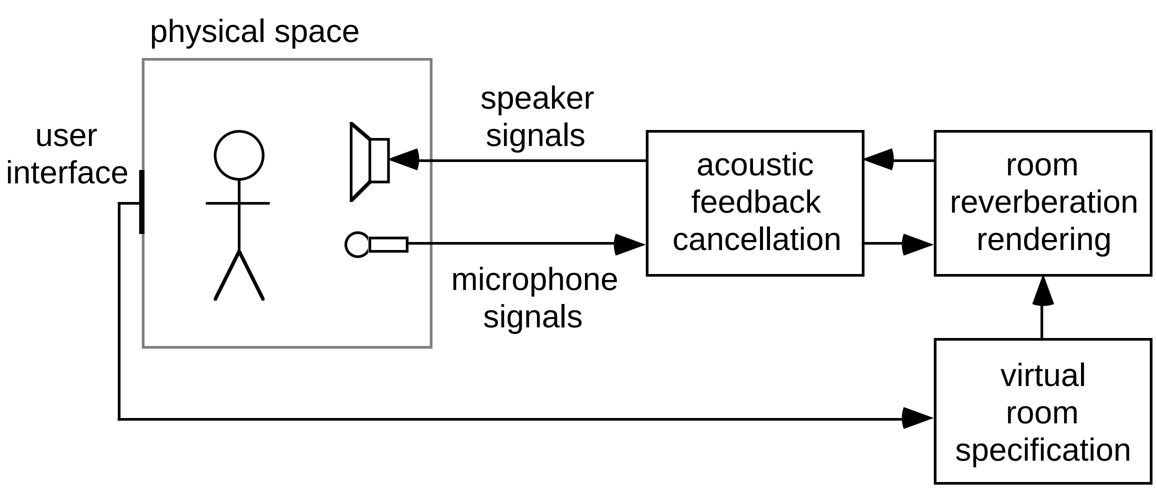


Figure 1.1 General block diagram of a virtual acoustic room.

1.3 Organization

Chapter 2, "Background" will review the most important developments in the disciplines related to this subject. The chapter starts with a review of room reverberation, and then continues with descriptions of reverberation enhancement systems, reverberation algorithms, room simulation, and echo cancellation.

Chapter 3, "Acoustic Feedback Cancellation" covers the topic of acoustic feedback cancellation via system measurement and linear filtering. The technique and its limitations are discussed and a simple theory is introduced. The amount of cancellation required to implement a virtual acoustic room is derived. Experimental results are given from both non-realtime simulations and a realtime cancellation system that was successfully developed.

Chapter 4, "Room Reverberation Modeling" covers the topic of simulating room reverberation with a small number of speakers located at the perimeter of the physical space. The problem is segregated into two parts: rendering the early echo response of a room using tapped delay lines based on the source image model, and rendering the later diffuse reverberant field using nested allpass reverberators. The reverberation algorithms are discussed in some detail. A four channel realtime experimental setup is described that successfully simulated a variety of different rooms.

Chapter 5, "Conclusions and Future Work" is a critical assessment of the work done and the results achieved. Areas of further research are indicated.

2. Background

2.1 Room Reverberation

The process of reverberation starts with the production of sound at a point within a room. The acoustic pressure wave expands radially outward, reaching walls and other surfaces where energy is both absorbed and reflected. Reflection off large, uniform, rigid surfaces produces a reflection the way a mirror reflects light, but reflection off non-uniform surfaces is a complicated process, generally leading to a diffusion of the sound in various directions. The wave propagation continues indefinitely, but for practical purposes we can consider the propagation to end when the intensity of the wavefront falls below the intensity of the ambient noise level.

Assuming a direct path exists between the source and the listener, the listener will first hear the direct response, followed by the reflections of the sound off large nearby surfaces, the so called early echoes. After a short period (one tenth of a second for typical rooms), the density of reflected waves becomes too high for discrete recognition. The remainder of the reverberant decay is characterized by a dense collection of echoes traveling in all directions, whose intensity is relatively independent of location within the room. This is called diffuse reverberation. The diffuse reverberation of good sounding concert halls decays exponentially [Schroeder-62]. The time required for the reverberation level to decay to 60 dB below the initial level is defined as the reverberation time.

Reverberation models treat the direct response and early echoes separately from the later diffuse reverberation. The direct response and early echoes consist of discrete wavefronts; they are directional and elicit a correlated response in the ears of the listener. They are also completely dependent on the orientation of the source, listener, and the major reflective surfaces. Consequently, the pattern and directionality of the early echoes provides the listener with information regarding the geometry of the physical space [Benade-85].

In contrast, the diffuse reverberation contributes to the spaciousness and timbre of the room, evoking a less specific response. Research in concert hall acoustics has confirmed that listeners respond favorably to lateral (left-right) reverberant energy, which results in uncorrelated signals at the two ears [Schroeder-74], [Barron-81]. Front-back reverberant energy would arrive simultaneously at the two ears, and would thus be correlated. The lack of binaural coherence is a main contributor to the perception of spaciousness of a room.

2.2 Reverberation Enhancement Systems

The largest amount of related work is found in the field of electroacoustical reverberation enhancement systems for concert halls. Reverberation enhancement systems address the frequently occurring problem of a concert hall that has poor acoustics. Generally, enhancement systems seek to increase the reverberation time of the hall at various frequencies. There have been several different approaches taken in the design of reverberation enhancement systems. I will briefly describe three relevant methods below.

A multichannel reverberation (MCR) system works by equipping a concert hall with many speaker-microphone pairs, each tuned to a specific narrowband frequency range [Parkin-65]. By increasing the loop gain for a speaker-microphone pair such that the acoustic feedback causes ringing, it is possible to arbitrarily lengthen the reverberation time at that frequency. An MCR system relies on acoustic feedback to lengthen the reverberation time, but the feedback is carefully controlled in each band. Essentially, this amounts to adding electroacoustic resonators to a room in order to increase the reverberation. An MCR system has no method of controlling early echo response.

Berkhout's Acoustic Control System (ACS) is based on the principle of reconstructed wavefronts [Berkhout-88]. Acoustic wavefronts produced by performers on stage are captured using a large array of microphones and reproduced for the audience using a large array of loudspeakers. An intermediate signal processing system allows the early echo response and diffuse reverberation of the room to be modified. Acoustic feedback is reduced by 1) directing the loudspeakers towards the audience and away from the microphones, 2) the use of directional microphones placed close to the sound source, and 3) the use of time variant reverberators. The ACS is essentially a multiple channel sound reinforcement system that relies on close miking to avoid feedback. As currently formulated, it is only applicable to situations involving a performance stage and an audience.

A recently developed method of reverberation enhancement requires the use of time variant reverberators [Griesinger-91]. A time variant reverberator can be constructed by inserting one or more time varying elements (delays or gains) into a standard reverberator. The resulting time variant system inhibits the buildup of acoustic feedback by constantly varying the phase response at all frequencies. Care must be taken to ensure that the time variation does not lead to perceptible frequency or amplitude modulation. Sounds created in the hall are picked up by a few microphones and passed through many time variant reverberators, each attached to a separate speaker bank. The use of many time variant channels increases the randomization effect of the reverberators and allows higher gain operation as well as distant mike placement. The result is a high gain, but stable, enhancement system. However, the use of time varying reverberators to achieve stability does not address the problem of the system hearing itself. In order to create arbitrary virtual spaces, acoustic feedback must be blocked. This is especially important for proper rendering of early echo response. Nevertheless, time varying reverberators can be used in conjunction with other feedback cancellation technology in order to enhance system stability without added coloration.

2.3 Reverberation Algorithms

Early efforts at simulating reverberation concentrated on the design of digital filters to mimic the response of rooms. These efforts were guided by the idea that the perceptual difference between a real room and a greatly simplified simulation could be made small [Schroeder-62]. Schroeder's initial reverberator design was composed of two types of infinite impulse response (IIR) filters, comb filters and allpass filters. A comb filter is a simple delay with feedback:

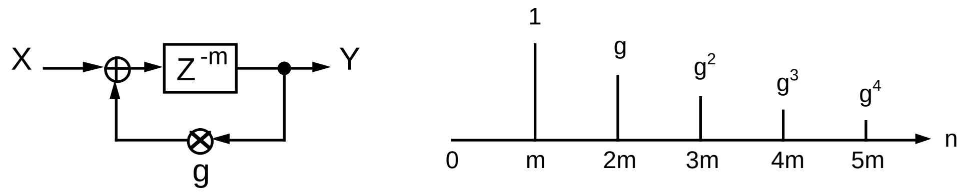


Figure 2.1 Comb filter flow diagram and impulse response.

The time domain impulse response of a comb filter is an exponentially decaying pulse train. Thus, the comb filter's response is somewhat analogous to an acoustic plane wave reflecting back and forth between two parallel walls. The pole-zero diagram of a comb filter contains poles evenly spaced around the unit circle with angles corresponding to the m^{th} roots of unity, and with magnitudes of the m^{th} root of g. The frequency response of the comb filter is a maximum at each pole location, and a minimum between poles (the resulting comb-like shape of the response is responsible for the name).

An allpass filter is like a comb filter with a feedforward path around the delay:

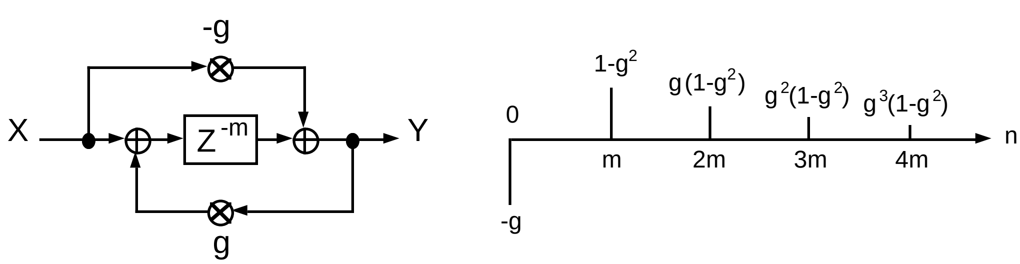


Figure 2.2 Allpass filter flow diagram and impulse response.

The pole-zero diagram of an allpass filter has the same pole configuration as the comb filter, but now for every pole, there is a zero at its conjugate reciprocal location. The zeroes cancel the influence of the poles on the magnitude of the frequency response, resulting in a flat frequency response. However, allpass filters do affect the phase of signals.

Schroeder's reverberator design consisted of four comb filters feeding into two serial allpass filters as shown in the figure below.

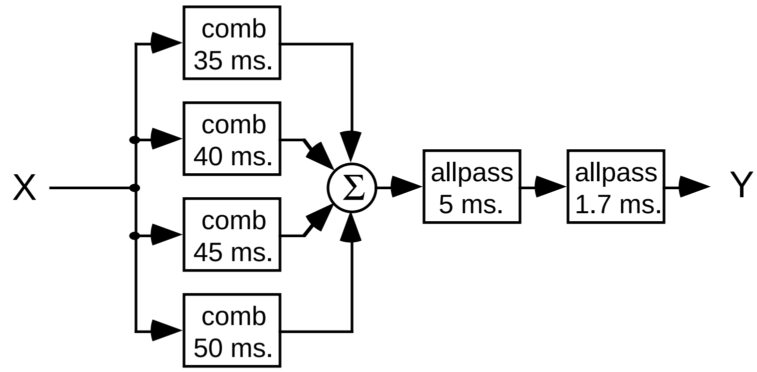


Figure 2.3 Flow diagram of Schroeder reverberator. Delay lengths shown in milliseconds, allpass filter gains are 0.7, comb gains are determined by desired reverberation time.

The basic idea of the parallel comb filters is to simulate the echoes that occur between walls in a concert hall; in the frequency domain, the peaks caused by the comb filters correspond to the normal modes of the hall. However, the parallel comb filters do not supply a sufficient buildup of echoes for realistic diffuse reverberation (in fact, the filters have a constant echo rate, so there is no buildup at all). In order to increase the echo density, the output of the comb filters is fed into one or more allpass filters in series. Each allpass filter has a multiplicative effect on the number of echoes, but prevents coloration because of the allpass filter's flat frequency response.

Although it was a breakthrough in its time, the Schroeder reverberator is quite poor by today's standards. The echo density does not build up sufficiently, and the reverberator has a very poor response to impulsive sounds, which create a rough, fluttering decay. Moorer revisited this basic design and made many improvements [Moorer-79]. More comb filters were added to achieve a greater echo density, the comb filters incorporated lowpass filters in their feedback loops to simulate frequency dependent air absorption, and the early echo response of the room was simulated by a tapped delay line. The Moorer reverberator does sound much more realistic than the Schroeder reverberator, but still exhibits a poor response to impulsive sounds.

2.4 Room Simulation

Unlike the study of reverberation algorithms, which are crude but efficient simulations of rooms, the study of room simulation is not concerned with efficiency, but accuracy. In general, these systems work by converting a detailed description of the room to be simulated (including the source and listener positions) into a binaural impulse response which is rendered by performing large convolutions with the input signal [Kleiner-91]. Methods for deriving the impulse response of the room include the source image method and acoustic ray tracing [Borish-84a].

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 identifies all virtual source positions out to a specified maximum distance. The free path propagation from these virtual sources to the listener position then determines the echo response. The figure below shows 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-84b].

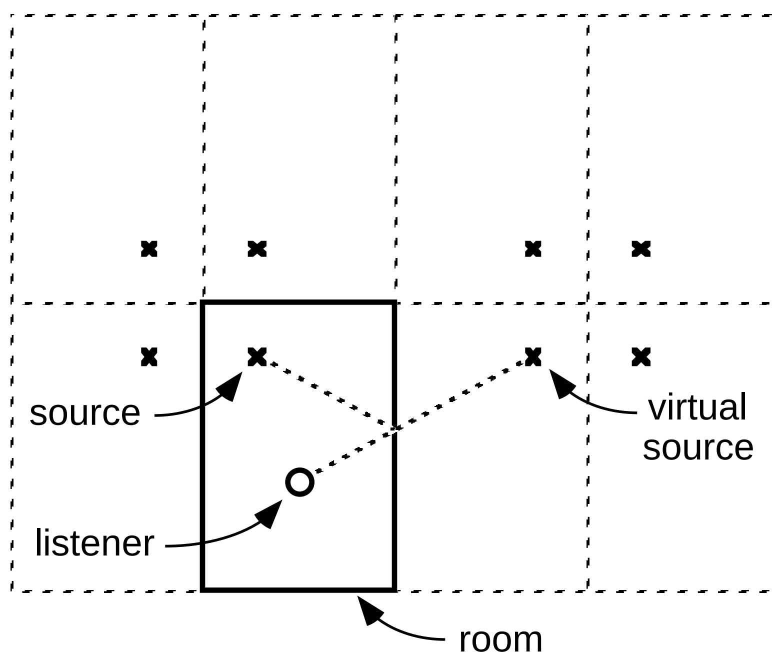


Figure 2.4 Virtual sources in a rectangular room. The dotted line from the source to the listener represents a reflected sound path which is equivalent to the free field contribution from the indicated virtual source. Additional virtual sources are shown that correspond to other reflective paths between the source and listener.

The ray tracing method works by propagating a large number of rays in all directions from the source position. Ray propagation continues linearly, reflecting off intersected walls, until the ray passes through a region close to the listener. This corresponds to a contribution to the listener (via the reflected path) from the source. By applying a statistical scattering to each reflection, diffuse reflective surfaces can be modeled. This makes ray tracing an excellent method for determining long term statistical properties of room reverberation [Schroeder-70]. For early echo determination, however, the source image method is a far more direct approach.

Using either method, a finite impulse response (FIR) filter is created from the composite contributions of each virtual source to the listener. The FIR tap delay lengths correspond to the sound travel time between the virtual source and the listener. The FIR coefficient amplitudes are proportional to the reciprocal of distance to the virtual source. All reflections also reduce the amplitude of the virtual source by a factor of the reflection coefficient of the wall material. A further improvement is to model the frequency dependence of air absorption and surface reflections. This is perfectly feasible, but increases the filter complexity.

Non-interactive room simulation has received considerable attention. The system described in the Kleiner reference accepts reverberant-free source material which is injected into a simulated room at a specified position. The binaural output corresponding to a specified listening position is computed and presented to a listener. Rather than use headphones, the binaural signal is delivered to the listener's ears using stereo speakers and a head related crosstalk cancellation filter [Schroeder-63]. The listener's position must remain fixed. Although the system is realtime, the researchers have accepted the constraints of non-interactivity and fixed listener position in order to improve the accuracy of the simulation.

Instead of simulating synthetic rooms, it is possible to record the binaural impulse responses of actual rooms [Schroeder-74]. The recorded responses can be used in place of the synthetic responses described above. In this way, the acoustics of different real rooms can be readily compared using the same source material.

Several room simulators have been developed that render the simulated reverberant field using large arrays of loudspeakers surrounding a listener in an anechoic chamber [Meyer-65] [Kleiner-81]. The systems described in the references utilized 65 and 52 loudspeakers, respectively. The 52 loudspeaker system simulated the early echo response using a 14 channel digital delay line (hence, 14 early echoes could be simulated), and the diffuse reverberant response was simulated using a reverberation chamber that provided four incoherent output channels.

2.5 Echo Cancellation

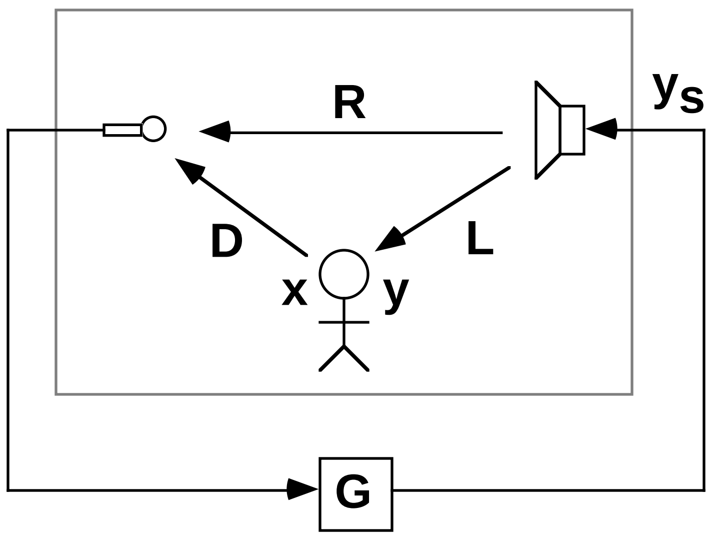
The problem of echo cancellation is a recurring one in speech recognition and telephony and is related to the problem of acoustic feedback cancellation. In speech recognition, the original speech signal is often accompanied by the reverberant response of the room, and it is desired to remove the reverberant echoes to recover the original signal. This is a problem of deconvolving two unknown signals [Oppenheim-89]. If by some chance the room response is known, then recovering the original speech is a matter of inverse filtering [Neely-79]; when the response is not known, a technique such as homomorphic deconvolution can perhaps be used [Schafer-68] [Stockham-75].

A more relevant application of echo cancellation occurs in telephony, where echoes are created when a 4-wire long distance link is attached to a 2-wire local link. The echo response is not known a priori, and is subject to change over time, but the transmitted speech signal is known; the problem is simply to estimate the echo response and use this to cancel the generated echoes. This problem has been solved using adaptive finite impulse response (FIR) filters [Sondhi-67]. Adaptive FIR filters have also been applied to the problem of acoustic echo cancellation in teleconferencing [Sondhi-91].

3. Acoustic Feedback Cancellation

3.1 Introduction

As discussed in the introduction, the problem of acoustic feedback between the speakers and microphones must be overcome in order to implement a virtual acoustic room. Let us review the system function of a generalized one channel sound reinforcement system:



x	sound source
y	final sound heard by listener
y_s	sound sent to speaker
D	transfer function from source to microphone
L	transfer function from speaker to listener
R	transfer function from speaker to microphone
G	transfer function of electronics (gain, reverb, etc.)

Figure 3.1 Generalized one channel sound reinforcement system.

We can easily solve for the system response from source to speaker, and from source to listener:

$$\frac{y_s}{x} = \frac{GD}{1-GR} \tag{3.1}$$

$$\frac{y}{x} = \frac{LGD}{1-GR} \tag{3.2}$$

The condition for system stability, that the loop gain be less than unity, is simply

$$|GR| < 1 \tag{3.3}$$

for all frequencies. The overall system gain is determined by LGD. Typically, L, the response from speaker to listener, and R, the response from speaker to microphone, are both reverberant room responses that are not controllable. Sound engineers adjust D, the response from source to microphone, and G, the electronics, to obtain acceptable system gain without instability or coloration. D is altered by changing the microphone to source distance, thus changing the amount of direct sound energy that the microphone receives.

The problem with the virtual acoustic room is shared by all reverberation enhancement systems, namely it is not possible to move the microphone(s) close to the sound source. In the virtual acoustic room, we desire a perimeter system of speakers and microphones. This is a similar situation to many reverberation enhancement systems in concert halls, where the microphones and speakers are intentionally hidden from public view. Consequently, the microphones are far from the source (hence D is small), and any effort to increase the overall gain LGD by increasing G causes the frequency peaks in GR to approach unit magnitude, which leads to ringing or uncontrolled feedback.

As Griesinger points out, coloration due to feedback can be reduced by:

1) Moving the microphones closer to the source.
2) Reducing the system level by reducing the system gain.
3) Increasing the number of independent channels.
4) Adding some form of time variation.

Increasing the number of independent channels is the approach taken in the multiple channel reverberation enhancement systems, and adding time variance is Griesinger's approach. I will investigate another possibility, that of directly canceling the feedback path R by measuring R and approximating it using an FIR filter. I call this predictive feedback cancellation.

3.2 Predictive Feedback Cancellation

The diagram below shows a one channel sound reinforcement system with a feedback cancellation filter:

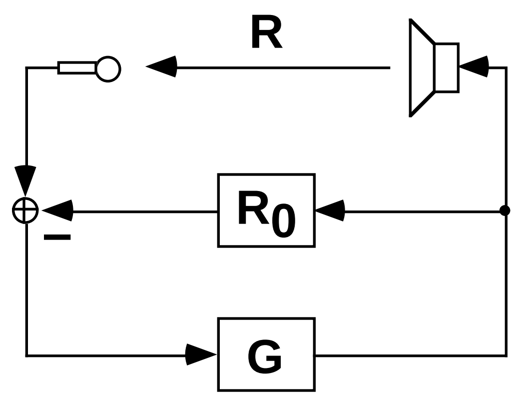


Figure 3.2 Predictive feedback cancellation.

In the above diagram, R is the system response between the speaker and microphone, and R_0 is an FIR filter that approximates R, obtained by measuring R directly. The loop gain of this system is

$$\frac{G(R-R_0)}{1-G(R-R_0)} \tag{3.4}$$

Clearly, as R_0 approaches R, the loop gain of this system approaches 0, indicating complete feedback cancellation. The success of this technique depends on various conditions:

1) The degree to which the system function R can be modeled as a linear, time-invariant system. We would expect R to be fairly linear, since speakers, microphones, amplifiers, D/A and A/D converters, and rooms can all be modeled well as linear systems. However, the system will not be entirely time-invariant, due to people moving about in the space, air currents caused by convection and ventilation, and changing atmospheric conditions.

2) The accuracy of the measurement of R. Presumably, the measurement is done once before each use. Assuming that the system is entirely linear and time-invariant, then the issues here are noise immunity and efficiency. Clicks, chirps, and noise bursts are commonly used measurement signals.

3) The complexity of R. Since R is essentially a room response, the complexity of R is determined by the reverberation time of the room. In the case of the virtual acoustic room, we have already mentioned that we intend to use an acoustically dead physical space, and thus R will have a small time support, perhaps a few hundred milliseconds.

4) The computational power allocated to implement R_0. Currently, commercially available processors can implement 128,000 point FIR filters in realtime, although they are expensive. On the other hand, an inexpensive signal processor can implement a 200 point FIR filter in realtime. These figures are for typical high quality audio sampling rates (44.1 kHz).

Another issue related to this technique is the number of cancellation filters needed for multiple channel systems. Each speaker microphone pair requires a cancellation filter, and thus the number of cancellation filters is equal to the product of the number of speakers and microphones.

3.3 Theory

The amount of cancellation can be defined as the ratio of the signal energy output from the microphone to the energy of this signal after the subtraction of the cancellation filter output. We would like to develop a simple theory to predict the amount of cancellation in decibels given the reverberation time of the physical space and the length of the FIR filter R_0. We will assume an ideal situation, where the system R is completely linear and time-invariant, and the measurement technique recovers R without degradation. After measuring, we create the FIR filter R_0 of length t_0 by choosing a rectangular window over R that maximizes the energy of the filter coefficients. Because R is a time-decaying room response, the filter coefficients will generally be taken from the start of R, as shown in the figure below.

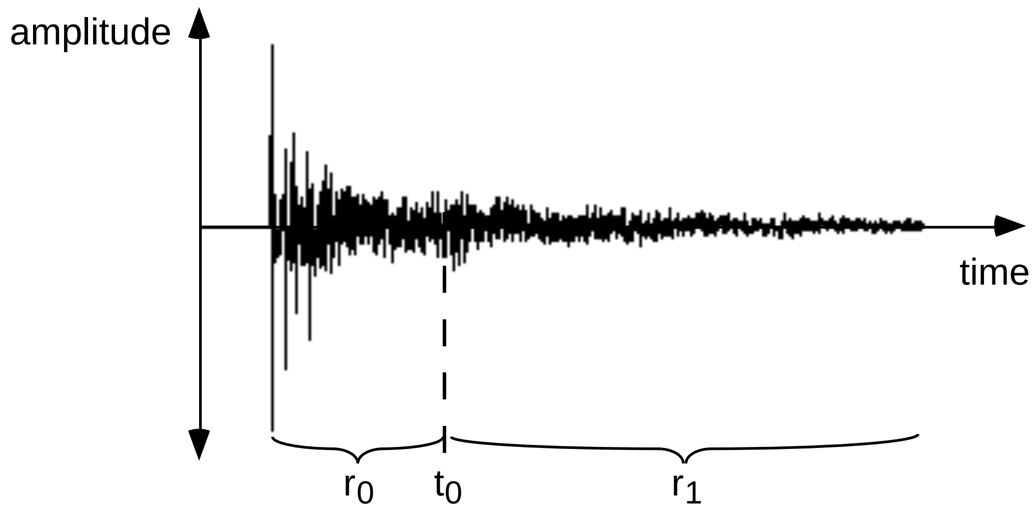


Figure 3.3 Impulse response R of typical room. Initial part of impulse response becomes the FIR cancellation filter (r_0). Remainder of response (r_1) is uncanceled.

Now we ask, how does the amount of cancellation depend on the reverberation time of R, the filter length t_0, and the input signal? Consider the following reorganized system:

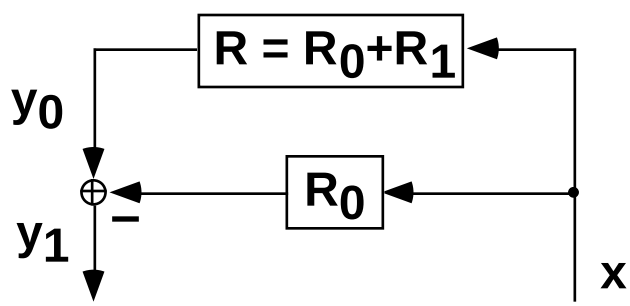


Figure 3.4 Predictive feedback cancellation.

Here, we consider the room response R as composed of R_0 (the portion being canceled) and R_1 (the portion not being canceled). The cancellation amount in decibels is

$$\text{Cancellation in dB} = 10\log_{10}\left(\frac{\text{Energy in } y_0}{\text{Energy in } y_1}\right) \tag{3.5}$$

where

$$y_o[n] = x[n] * (r_0[n] + r_1[n]) \tag{3.6}$$
$$y_1[n] = x[n] * r_1[n] \tag{3.7}$$

Expanding the inner ratio in equation 3.5:

$$\frac{\text{Energy in } y_0}{\text{Energy in } y_1} = \frac{\sum_n \left(x[n] * r_0[n] + x[n] * r_1[n]\right)^2}{\sum_n \left(x[n] * r_1[n]\right)^2} \tag{3.8}$$

$$= \frac{\sum_n \left(x[n] * r_0[n]\right)^2 + 2\left(x[n] * r_0[n]\right)\left(x[n] * r_1[n]\right) + \left(x[n] * r_1[n]\right)^2}{\sum_n \left(x[n] * r_1[n]\right)^2} \tag{3.9}$$

Using Parseval's theorem:

$$= \frac{\int_{-\pi}^{\pi} |\mathrm{X}(\omega)|^2 |R_0(\omega)|^2 \, d\omega + 2\int_{-\pi}^{\pi} |\mathrm{X}(\omega)|^2 R_0(\omega) R_1^*(\omega) d\omega + \int_{-\pi}^{\pi} |\mathrm{X}(\omega)|^2 |R_1(\omega)|^2 \, d\omega}{\int_{-\pi}^{\pi} |\mathrm{X}(\omega)|^2 |R_1(\omega)|^2 \, d\omega} \tag{3.10}$$

To simplify, let's assume x[n] is a broadband signal such that

$$|X(\omega)| = 1 \quad \text{for all } \omega \tag{3.11}$$

Furthermore, the cross product terms in equation 3.10 can be eliminated noting that r0[n] and r1[n] are nonzero over different ranges of n and so their product is zero for all n:

$$\sum_n r_0[n] r_1[n] = 0 \quad \xleftrightarrow{f} \quad \int_{-\pi}^{\pi} R_0(\omega) R_1^*(\omega) d\omega = 0 \tag{3.12}$$

Thus, the energy ratio simplifies to

$$\frac{\text{Energy in } y_0}{\text{Energy in } y_1} = \frac{\int_{-\pi}^{\pi} |R_0(\omega)|^2 \, d\omega + \int_{-\pi}^{\pi} |R_1(\omega)|^2 \, d\omega}{\int_{-\pi}^{\pi} |R_1(\omega)|^2 \, d\omega} \tag{3.13}$$

$$= \frac{\text{Energy in } r_0 + \text{Energy in } r_1}{\text{Energy in } r_1} \tag{3.14}$$

Thus, for simple broadband input signals such as an impulse or white noise, the amount of cancellation is determined by the ratio of energy in the total response r to the energy in the uncanceled portion r_1. Note that for narrowband signals, the cancellation filter can have an undesired effect. Imagine a multipath propagation from a speaker to a microphone such that the microphone is at a pressure node for some frequency. Applying a cancellation filter that cancels some, but not all, reflected paths will increase the canceled (y_1) signal strength at that frequency. Even for broadband input signals, some frequencies will be boosted by the cancellation filter, but the average effect is to attenuate signal energy.

3.4 Energy Decay of Room

To relate the findings of the previous section to real rooms, we consider the system function R to be a room's impulse response which is modeled as a broadband noise signal with an exponentially decaying envelope. The envelope of the response is depicted below:

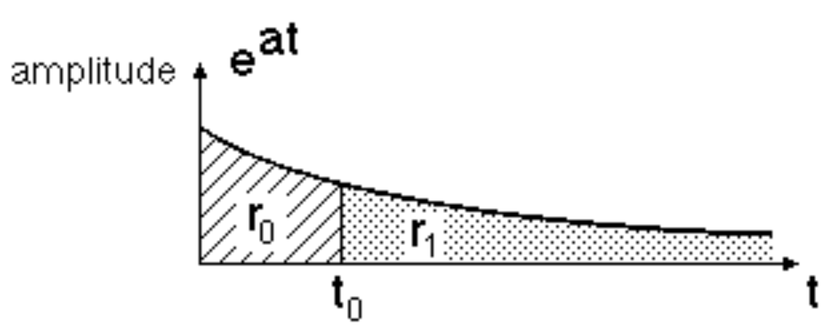


Figure 3.5 Amplitude envelope of idealized room response.

In this diagram, t_0 is the length of the FIR cancellation filter, which partitions the response into the r_0 and r_1 sections. To determine the cancellation, we are interested in the ratio of energies as given by equation (3.14). This ratio is expressed below:

$$\frac{\int_0^\infty \left(e^{at}\right)^2 dt}{\int_{t_0}^\infty \left(e^{at}\right)^2 dt} = e^{-2at_0} \tag{3.15}$$

The relationship between the decay factor (a) and the 60 dB reverberation time (T) is simply:

$$e^{aT} = \frac{1}{1000}, \quad a = \frac{\ln(0.001)}{T} = \frac{-6.91}{T} \tag{3.16}$$

Combining equations 3.5, 3.15, and 3.16, the cancellation in decibels given the reverberation time of the physical space (T) and the length of the FIR filter (t0) is:

$$\text{Cancellation in dB} = 10\log_{10}\left(e^{13.82t_0/T}\right) = 60\frac{t_0}{T} \tag{3.17}$$

which is hardly a surprising result. This simply says that if we choose t_0 to be equal to the 60 dB reverberation time of the room, then we would expect 60 dB of cancellation. This result, although correct for highly idealized simulations, is inadequate for predicting actual cancellation amounts in real rooms, as we will see in the experimental section.

3.5 Required Cancellation for a Virtual Acoustic Room.

How much cancellation is required to implement a virtual acoustic room? Griesinger's approach to solving this problem is to relate the loop gain of the reverberation enhancement system to the enhanced critical distance of the virtual acoustic space [Griesinger-91], and I will draw heavily from his paper in this section.

The critical distance of a room is the distance from a sound source at which the direct sound level is equal to the reverberant sound level. Thus, the critical distance is a measure of the reverberation level of a room. Typical concert halls have critical distances of about 7 meters; a living room might have a critical distance of under 1 meter. When a reverberation enhancement system is active, the reverberation level in the room will generally increase, thus establishing an enhanced critical distance that is smaller than the natural critical distance of the room.

Considering a system with a single speaker and microphone, the average loop gain of the system is the average microphone output from the speaker divided by the average microphone output from the source:

$$\text{Avg. loop gain} = \frac{\text{avg. mike output from speaker}}{\text{avg. mike output from source}} \tag{3.18}$$

Referring to figure 3.1, the average loop gain is:

$$\text{Avg. loop gain} = \frac{Ry_s}{Dx} = \frac{GR}{1-GR} \tag{3.19}$$

Quoting directly from Griesinger,

> *"In a broadband system the loop gain is an average over many frequencies. The transfer function between the speaker and microphone has many peaks and valleys as a function of frequency due to the interference between the many reflections in the sound path. The loop gain at some frequencies is much higher than the average. As the gain in the system is increased the system rings at the frequency of the highest peak. If we assume the microphone and the loudspeaker are separated by at least the critical distance of the room, the average loop gain where ringing begins has been predicted by Schroeder* (see [Schroeder-54]). *The maximum gain depends on the reverberation time of the room and the bandwidth of the system, and is always much less than unity. For a broadband system and a reverberation time of two seconds the maximum loop gain is about -12 dB. In addition, to avoid*

obvious coloration in a broadband system the loop gain should be at least 8 dB less than the gain at which ringing begins. This means for a high quality reinforcement or acoustic enhancement system the average loop gain must be -20 dB or less!"

For the case of a sound reinforcement system, the response R in equation 3.19 will generally be a reverberant room response, and G will be a simple gain. However, in a virtual acoustic room, the response R will generally have significantly less reverberation than the response G, which is the synthesized reverberant response. Nevertheless, we need only consider the reverberation time of the product of these responses (GR), when determining the maximum loop gain.

We can easily reformulate the average loop gain based upon the distance from the source to the microphone and the enhanced critical distance as follows:

$$\text{Avg. loop gain} = \frac{\text{source distance}}{\text{enhanced critical distance}} \tag{3.20}$$

For the virtual acoustic room, we know the distance from the source to the microphone, as well as the enhanced critical distance of the synthesized virtual space. Using equation 3.20, we can then determine the average loop gain of the virtual acoustic room without acoustic feedback cancellation. In general, this loop gain will be much higher than the -20 dB maximum loop gain established for high quality reinforcement. The difference between the two will indicate the amount of cancellation required:

$$20\log_{10}(\text{avg. loop gain}) + \text{cancellation} \leq -20\ \text{dB} \tag{3.21}$$

where the cancellation is expressed in decibels. The enhanced critical distance of the virtual acoustic space depends on the reverberation time and volume of the virtual room. From Kuttruff, the relationship given for real rooms is:

$$r_h = 0.1\left(\frac{V}{\pi T}\right)^{1/2} \tag{3.22}$$

where r_h is the critical distance of the room, V is the volume of the room in cubic meters, and T is the 60 dB reverberation time [Kuttruff-91]. Note that this will apply equally to a virtual acoustic room provided that the rendering of the virtual room conserves energy, and therefore changing the virtual room volume or reverberation time will affect the enhanced critical distance as in equation 3.22. Combining equations 3.21 and 3.22, the cancellation required is:

$$\text{Cancellation in dB} \leq -20\left(1+\log_{10}\left(\frac{\text{source distance}}{0.1\left(V/\pi T\right)^{1/2}}\right)\right) \tag{3.23}$$

This result assumes that the microphone is omnidirectional and is placed such that it receives the same reverberation level as a listener. Less cancellation is required if a directional microphone is used or if the speaker placement directs most of the synthetic reverberant energy toward the listener rather than the microphone (however, because small speakers radiate omnidirectionally at low frequencies, it is difficult to affect low frequency feedback through speaker placement).

An example of the use of equation 3.23 follows. Imagine we are simulating Boston Symphony Hall in a physical space with a source to microphone distance of 2 meters. Symphony Hall has a volume of 18,800 cubic meters and a reverberation time of 1.8 seconds (500-1000 Hz). This yields a critical distance of 5.8 meters, and thus -10.8 dB of cancellation is required.

Unfortunately, the cancellation required as given by equation 3.23 cannot simply be equated with the cancellation in equation 3.17 (the amount of cancellation obtained as a function of FIR length and reverberation time). This is because the loop gain analysis requires that the gain be decreased at all frequencies so that peaks in the spectrum do not ring. Equation 3.17 gives an average cancellation, but says nothing about peaks in the canceled spectrum.

3.6 Cancellation Experiments

The remaining sections in this chapter will discuss various experiments conducted regarding predictive feedback cancellation in rooms. The experiments were designed to determine the amount of feedback cancellation possible in a real situation. Specifically, the experiments involved cancellation between a single speaker microphone pair in various sized rooms. The diagram below shows the complete system that was to be measured and predicted:

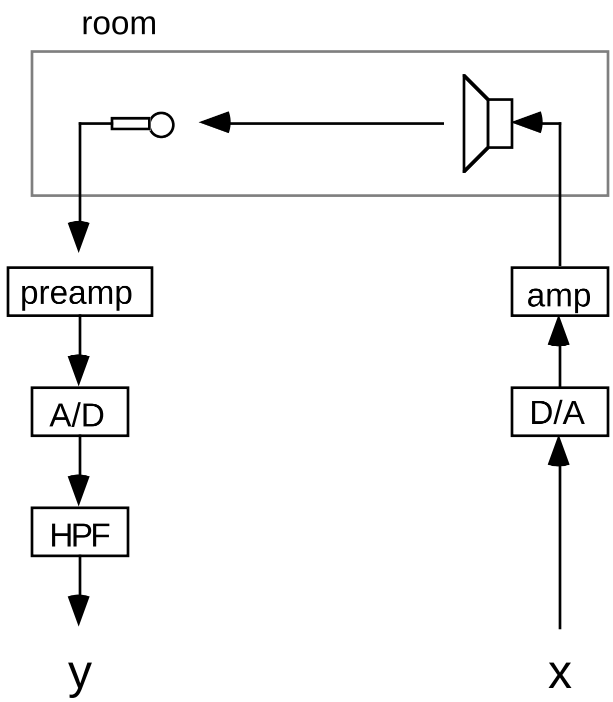


Figure 3.6 Block diagram of experimental cancellation setup.

In the above diagram, x is the input to a system that consists of a D/A converter, a power amplifier, a speaker, room, and microphone, a microphone preamplifier, A/D converter, and a digital AC-coupling highpass filter. The AC-coupling filter is necessary because of the DC offset added by A/D converters which causes a non-linear transfer function. The filter used was a one pole, one zero implementation with a -3 dB corner frequency of 20 Hz.

The experiments consisted of the following steps:

1) Set up components, calibrate system gain, and measure ambient noise level.

2) Measure the impulse response of the system.

3) Play a noise burst and a speech signal through the system and record the results.

4) Using various length FIR filters to approximate the system response, determine how much energy cancellation is obtained for each signal.

5) Repeat with a different room.

The equipment used included an Apple Macintosh IIx computer with a Digidesign Audiomedia DSP card (containing 16-bit A/D and D/A converters), a Pioneer A-337 integrated amplifier, a Pioneer S-T1000 bookshelf type loudspeaker, a Neumann KM84i cardiod microphone, and a Yamaha MLA7 microphone preamplifier.

As seen from the DSP card, the system is a discrete time 16-bit digital system (at a 44.1 kHz sampling rate), and all energy measurements were done relative to a full scale 16-bit square wave, which is defined as an energy of 0 dB. The system gain was set by adjusting the amplifier output gain and microphone preamp gain so that the broadband system gain was approximately -4 dB, and so the components were operating in nominal, linear regions. The ambient noise level of the system was measured by recording 2 seconds of sound, calculating the recorded signal's energy, and averaging several such measurements.

The system response was measured using ML sequences (maximum length pseudo-random binary sequences; an excellent description of ML sequence measurements is given by Rife [Rife-87]). The amplitude of the measurement ML sequences was 8191, thus their energy was -12 dB. The period length of the ML sequence depended on the room being measured, and was either 65535 samples (1.5 seconds) or 32767 samples (0.74 seconds). The test signals were a 2 second gaussian noise burst with an energy of -12.3 dB, and a 2 second speech signal "this is a test of sampled speech" with an energy of -16.4 dB.

Three rooms were tested, the Experimental Media Facility (the “Cube”), my office (E15-491) and a small listening booth (E15-484a). The table below gives relevant data for the three rooms:

room	w•l•h (ft)	vol (m^3)	d_{src} (m)	T (sec)	noise (dB)	gain (dB)	r_h (m)
Cube	60•60•50	5100	3.5	0.61	-48	-4	5.2
Office	8•13•12	35	2	0.34	-57	-5	0.56
484a	5•8•7	8	1	0.12	-66	-3	0.46

The table gives the dimensions and volume of the room, the speaker to microphone distance (d_{src}), the 60 dB reverberation time (T), ambient noise level, system gain, and critical distance (r_h) calculated using equation 3.22. The reverberation time was obtained directly from the energy contour of the system responses (using a 5 msec averaging window). Because the noise level of the system responses was generally -40 to -50 dB (somewhat higher than the ambient noise), the 60 dB reverberation time was estimated by linear extrapolation of the 20 and 40 dB reverberation times. Because of the multislope nature of room responses, this technique may yield shorter reverberation times than more classical techniques. The figure on the following page shows the impulse response and energy contour of my office.

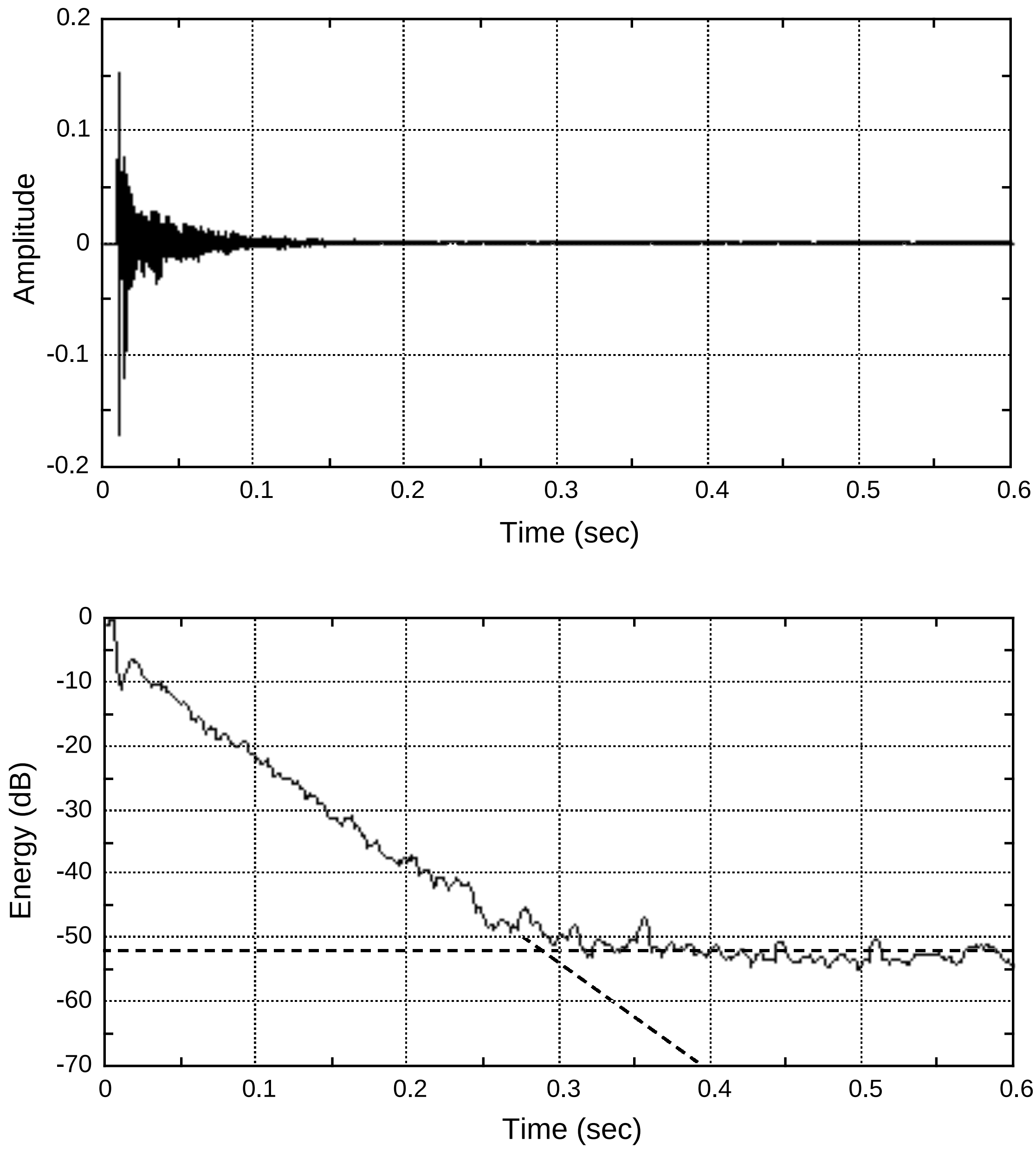


Figure 3.7 Measured impulse response of office (top). Energy contour of same impulse response using 5 millisecond averaging window (bottom). Dotted lines show noise level of response at -52 dB and extrapolation of energy decay.

The FIR cancellation filter lengths varied from 128 samples to 32K samples in powers of two. The FIR filter coefficients were obtained by sliding a rectangular window (whose length was the desired FIR length) along the system response and finding the window position that maximized the energy of the coefficients under the window. Then, the FIR cancellation filter was composed of the chosen coefficients in series with a delay corresponding to the start location of the window. In this manner, the FIR filter did not have to contain leading zeros corresponding to the sound travel time between speaker and microphone. The cancellation amount was calculated by comparing the energy of the recorded signal with the energy of the canceled signal. The canceled signal was obtained by convolving the original signal with the FIR filter response and subtracting the result (appropriately delayed) from the recorded signal (see figure 3.3). The following plots show the cancellation amount in decibels as a function of FIR filter length for both the noise and speech test signals:

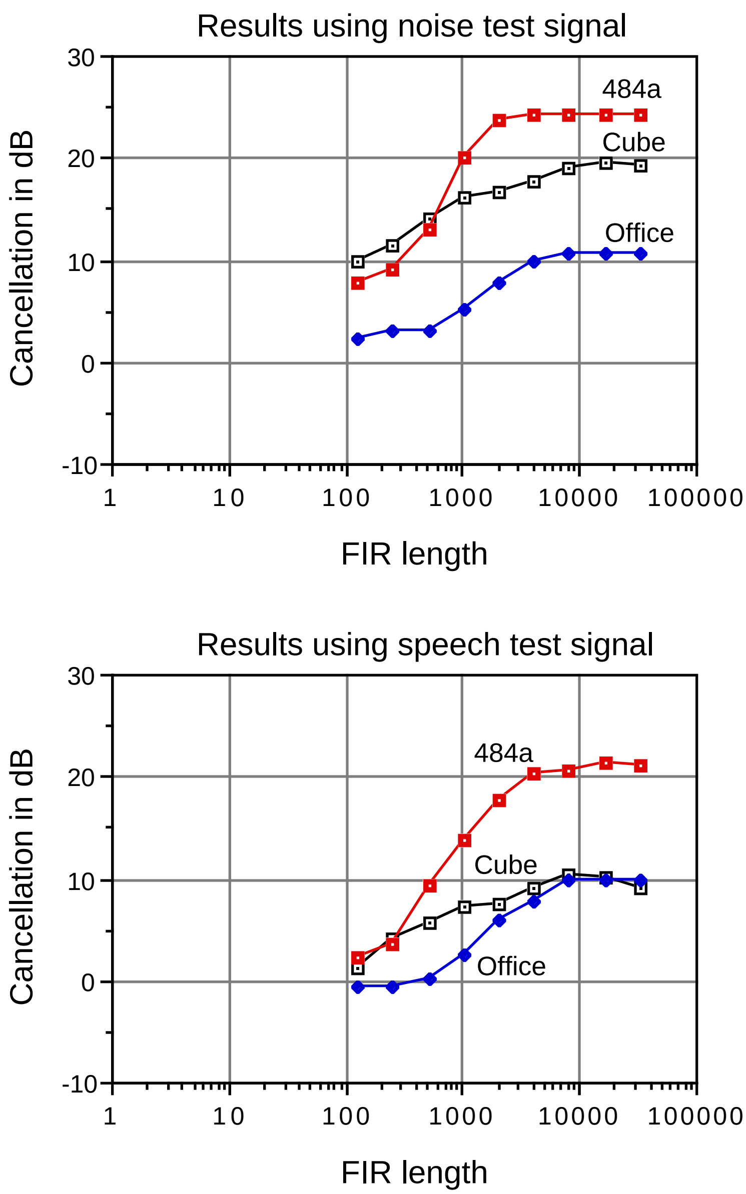


Figure 3.8 Cancellation results for noise and speech signals.

The noise test signal yielded better cancellation than the speech signal for all three rooms. In room 484a, the noise signal was attenuated by over 20 dB with a 1024 point FIR filter, which is an excellent result. For both the noise and speech signals, room 484a yielded the best results, followed by the Cube, and the office had fairly dismal results. In fact, the office results were so poor that the experiment was completely redone using different speaker and microphone positions, but the results were essentially the same. According to our primitive theory (equation 3.17), we would expect the cancellation to be best in rooms with small reverberation times, thus the office should have faired better than the Cube.

There are two possible reasons why the office results are so poor, relating to time variation and non-linearity in the room response. First, I was in my office during the measurement procedure, and even though I was fairly motionless during each measurement, I undoubtedly shifted positions between the ML sequence measurement and the playback/recording of the test signals. I was similarly present in the Cube during measurements, but the Cube has almost 150 times the volume of my office, so it seems unlikely that a small shift in my position would affect the Cube's response greatly. Room 484a was empty during measurements. It is possible that small changes in my position affected the room response significantly, or perhaps other factors are involved, such as air currents due to ventilation and convection. Another possibility is that the office has a fairly non-linear room response due to the presence of various objects that buzz and rattle non-linearly when excited acoustically. A room response that is time varying and non-linear adversely affects both the room response measurement and the subsequent cancellation filtering. ML sequence measurements recover the linear, time-invariant portion of the room response; any time variation or non-linearity will show up as noise in the measurement. This is why the noise level of each room response measurement was higher than the measured ambient noise level of the room. The cancellation filter effectively cancels the linear, time-invariant portion of the room response, leaving distortion products untouched. Listening to the canceled office speech signal revealed a distorted result, curiously sounding more like ring modulation than harmonic distortion.

The results of room 484a are closest to ideal for several reasons. The room was designed to be acoustically isolated and extremely dead; the walls are thick fabric covered fiberglass panels, the floor is carpeted, and the ceiling is acoustical tile. Measurements were made remotely with the speaker and microphone enclosed in the empty room. The plot below compares the calculated cancellation and the actual results. The calculated cancellation has been clipped to the maximum possible cancellation, which is simply the energy difference between the test noise signal and the ambient noise (in this case, 36.7 dB).

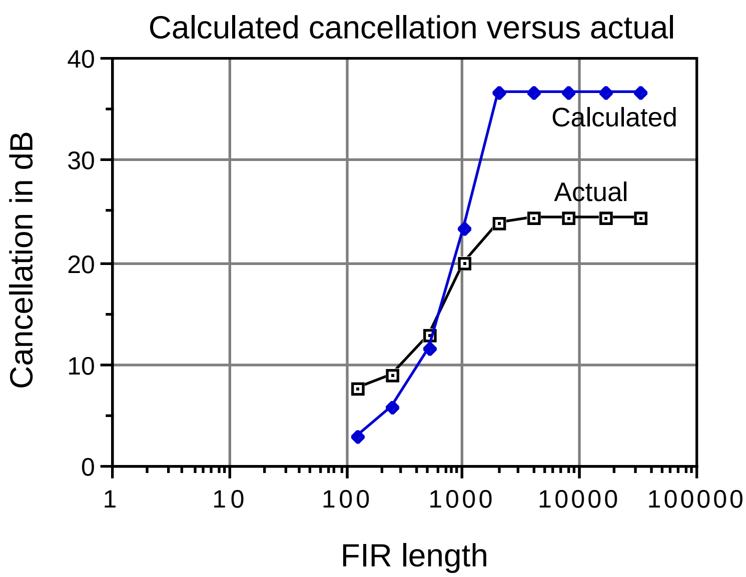


Figure 3.9 Calculated cancellation results and actual results.

The actual cancellation results show more cancellation for small FIR filter lengths. This is because the start of the system response has a significant amount of direct sound energy even though the microphone to speaker distance was greater than the critical distance of the room. For large FIR filter lengths, the actual cancellation reaches a limit of 24.3 dB. Listening to the canceled speech signal (17.8 dB cancellation with a 2048 point FIR filter), revealed an attenuated, but distorted result. Clearly, some element in the audio path (most likely the speaker) was distorting, and by removing the linear portion of the result via the cancellation filter, we are left with the remaining distortion products.

The cancellation model could be greatly improved by accounting for the following factors:

1) Ambient noise.
2) Non-linearity (distortion).
3) Time variation.
4) Speaker to microphone distance vs. critical distance (i.e. proportion of direct to reverberant sound).

Ambient noise, non-linearity, and time variation all contribute noise to the room response measurement and constrain the maximum cancellation possible. Ambient noise and non-linearity can be easily measured. Time variation could be measured by making many ML sequence measurements and determining the variance of the results. Any variation not accounted for by ambient noise or distortion would be considered a system response variation. Accounting for the proportion of direct to reverberant sound allows us to better predict the behavior of short FIR cancellation filters.

3.7 Realtime Cancellation

A realtime cancellation system was programmed using a Digidesign Audiomedia DSP card. The card contains a 20 MHz Motorola 56001 DSP, sample memory, and stereo, 16-bit A/D and D/A converters. The block diagram below shows the signal flow path of the realtime cancellation system:

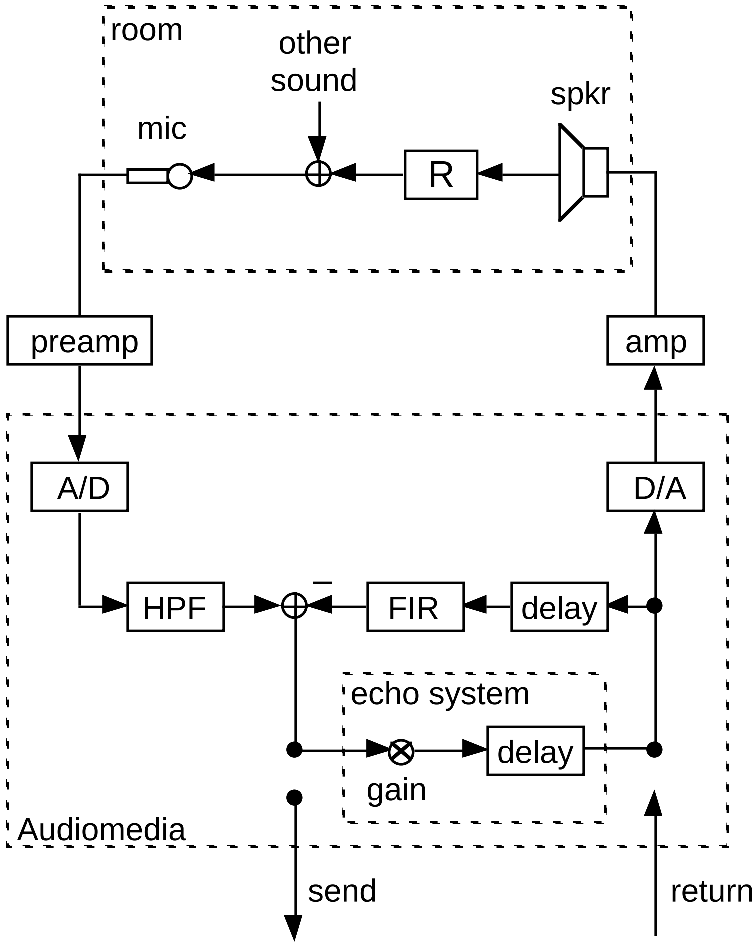


Figure 3.10 Realtime cancellation block diagram.

Before running the realtime cancellation software, the system response is first measured using an ML sequence. Then the realtime cancellation software is loaded into the Audiomedia along with the FIR coefficients and FIR delay. In the above diagram, any sound sent to the speaker is also sent through a delay and then the FIR cancellation filter. The implementation allows a 160 point FIR filter to be computed in realtime. The cancellation filter output is subtracted from the input signal which has passed through the AC-coupling highpass filter (HPF). The canceled result is sent to a simple echo system consisting of a programmable gain and delay, or it can optionally be sent to an external processor via an analog send/return loop.

I experimented with the realtime setup in the Cube. The speaker and microphone were placed relatively close together (about 1 meter). The 160 point filter typically yielded 10 dB of broadband cancellation. The cancellation filter could be enabled/disabled without affecting the ambience system, which allowed a convincing demonstration of the feedback cancellation. With a delay length of about 100 milliseconds, sounds picked up by the microphone were noticeably delayed and then echoed through the speaker. The acoustic feedback was attenuated by the cancellation filter such that only one echo was distinctly audible. Disabling the cancellation filter changed the system dramatically, causing a succession of echoes to be heard. If the gain was set high enough, the system could be switched from stable operation to unstable operation by enabling and disabling the cancellation filter. Inserting a reverberator into the send/return loop gave the expected result: when the reverberation time was long, say 2 seconds, the reverberation suffered from ringing and coloration until the system gain was reduced significantly, such that the resulting system was insensitive to all but very loud, impulsive sounds.

3.8 Conclusions

It is apparent that predictive cancellation using FIR filters can be effective at canceling acoustic feedback. This is true not only in an ideal case (room 484a), but also in a non-ideal case (the Cube). It remains to be determined why the cancellation faired so poorly in my office.

This research has not addressed the very important issue of how a room response changes over time due to people moving within the room. If small motions of people cause drastic changes in the room response, then it is doubtful that predictive cancellation will be of much use, unless adaptive techniques can successfully compensate for the changing room response. Room response changes due to ventilation, convection currents, and changing atmospheric conditions also need to be quantified.

Another shortcoming is the failure to satisfactorily bridge the two equations 3.17 (expected cancellation as a function of FIR length and reverberation time) and 3.23 (required cancellation). Clearly, the goal should be to determine a set of relationships that enable us to make design tradeoffs and predict the success or failure of an architecture.

4. Room Reverberation Modeling

4.1 Introduction

The goal of this chapter is to develop techniques for simulating room acoustics in the context of the virtual acoustic room. That is, instead of creating binaural output for listening through headphones, we desire a system that will render the simulated acoustics through an arbitrary number of loudspeakers located at the perimeter of the physical space. We will allow the listener to face any direction, but to make things easier we will assume the listener is near the center of the physical space. Clearly, more accurate simulations will be possible with more speakers, but implementation concerns dictate that we minimize the number of speakers, as well as the computational requirement per speaker. Thus, the task is to simulate the gross perceptual qualities of different acoustical spaces using a small number of speakers placed around a perimeter. Similar constraints are placed upon auditorium simulators for home use [Borish-85] [Griesinger-89]. In an interactive virtual acoustic room implementation, the source signal to the reverberation simulator will be taken from one or more microphones after the feedback cancellation processing has been done. However, to devise an adequate room simulator, it is sufficient to use any source material, such as a compact disc recording, although the material should be free from artificial reverberation.

We want the simulation to be driven from a specification of the desired virtual acoustic space. Accurate room simulation requires a detailed geometrical and material description of the room, but for our purposes a much simpler specification will do. Our specification will include the geometry of the perimeter of the virtual space, realized as a polyhedron, and a description of the broadband absorption coefficients of the planar wall surfaces. Also necessary is the location of the physical space within the virtual space. This specification will allow us to determine an early echo response for the room as well as calculate the reverberation time based upon the room's volume and absorption [Kuttruff-91] [Beranek-86] [Sabine-00]. Finally, given a description of the speaker positions along the perimeter of the physical space, a digital filter can be constructed for each speaker to render the simulated acoustical space. Success of the simulation can be determined by listening to source material rendered in the context of various simulated spaces.

Towards this end, a four channel audio system was constructed in the Cube to test the concepts described in this chapter. The layout of the audio system is shown below:

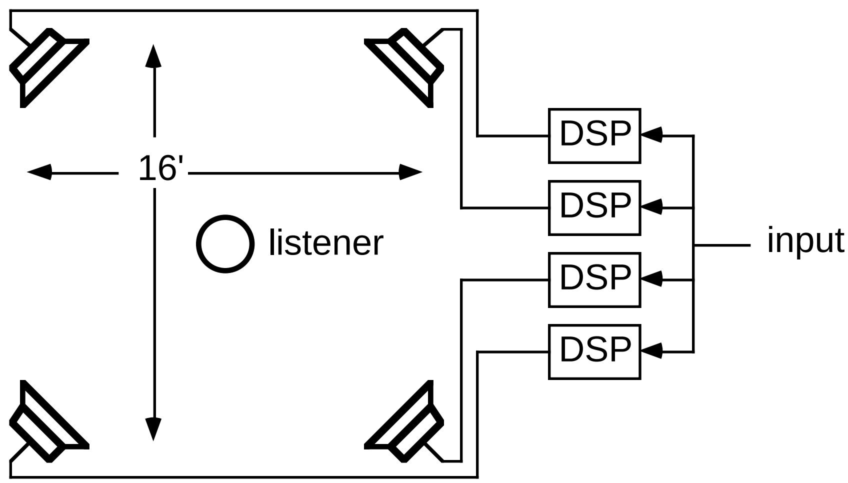


Figure 4.1 Four channel experimental audio system.

The four speakers were placed at the corners of a 16 foot square at a height of 5 feet. The listener sat (or stood) near the center of this square. Each of the four channels was driven by a separate Audiomedia DSP card, thus the computational engine to render the reverberation consisted of four 20 MHz. Motorola 56001 DSPs. The control program to run the DSPs is briefly described in a later section. Stereo or monophonic source material was supplied from a compact disc player.

As we will see in the next section, a four channel system is somewhat inadequate at rendering the proper spatial location cues because of the large angular spacing between adjacent speakers. Also, this system is restricted to rendering virtual sources in the horizontal plane, because of the lack of overhead speakers. However, this did not prove to be a serious limitation, as the diffuse soundfield from surround speakers is capable of delivering a very spacious effect.

4.2 Early Echo Rendering

A program was written that reads a room specification and the source and listener positions and calculates a set of virtual source positions using the source image method. The program was greatly simplified by only considering a two dimensional world consisting of the horizontal plane of the listener. Thus, floor and ceiling reflections were not considered in determining the early echo pattern. On the horizontal plane, all virtual sources were found within a specified distance from the listener. The list of virtual source positions was 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 [Moore-90]. 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 physical space and a listener at the center of the space:

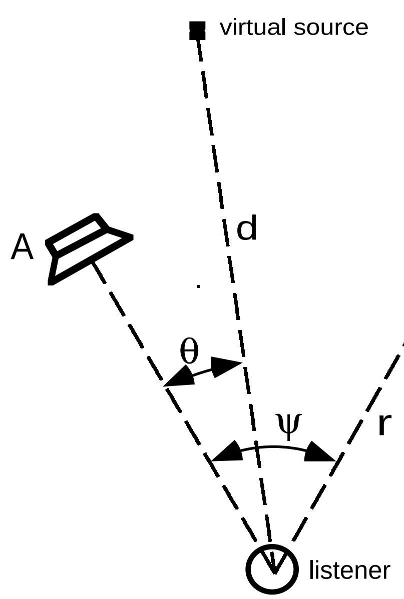


Figure 4.2 Intensity panning between adjacent speakers.

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:

$$\text{A,B tap delays} = \frac{d-r}{c} \tag{4.1}$$

$$\text{A tap amplitude} = a\frac{r}{d}\cos\left(\frac{\pi\theta}{2\psi}\right) \tag{4.2}$$

$$\text{B tap amplitude} = a\frac{r}{d}\sin\left(\frac{\pi\theta}{2\psi}\right) \tag{4.3}$$

$$a = \prod_{j\in S}\Gamma_j \tag{4.4}$$

where c is the speed of sound, a is the amplitude of the virtual source relative to the direct sound, S is the set of walls that the sound encounters, and Γ_j is the reflection coefficient of the j^{th} 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.

Early work in quadraphonic sound systems revealed deficiencies in the use of four speakers arranged in a square [Theile-77]. Referring to the diagram below, with the listener facing forward, it is difficult to render lateral phantom sources because small level differences between front and rear speakers cause large angle changes in the localization of the source.

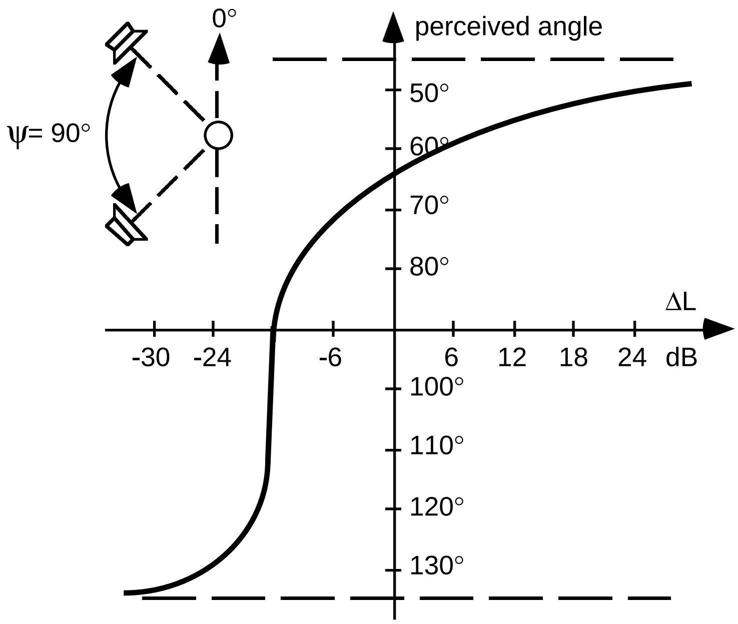


Figure 4.3 Direction of phantom source versus interchannel level difference for the lateral loudspeakers of a quadraphonic arrangement [Ratliff-74]. The listener is facing forward. Note the large change in perceived angle for very small level differences around 100°.

Theile determined that six loudspeakers arranged at 60 degree intervals was sufficient for proper localization of phantom sources using intensity panning. The diagram below shows the desired versus perceived sound direction using six loudspeakers:

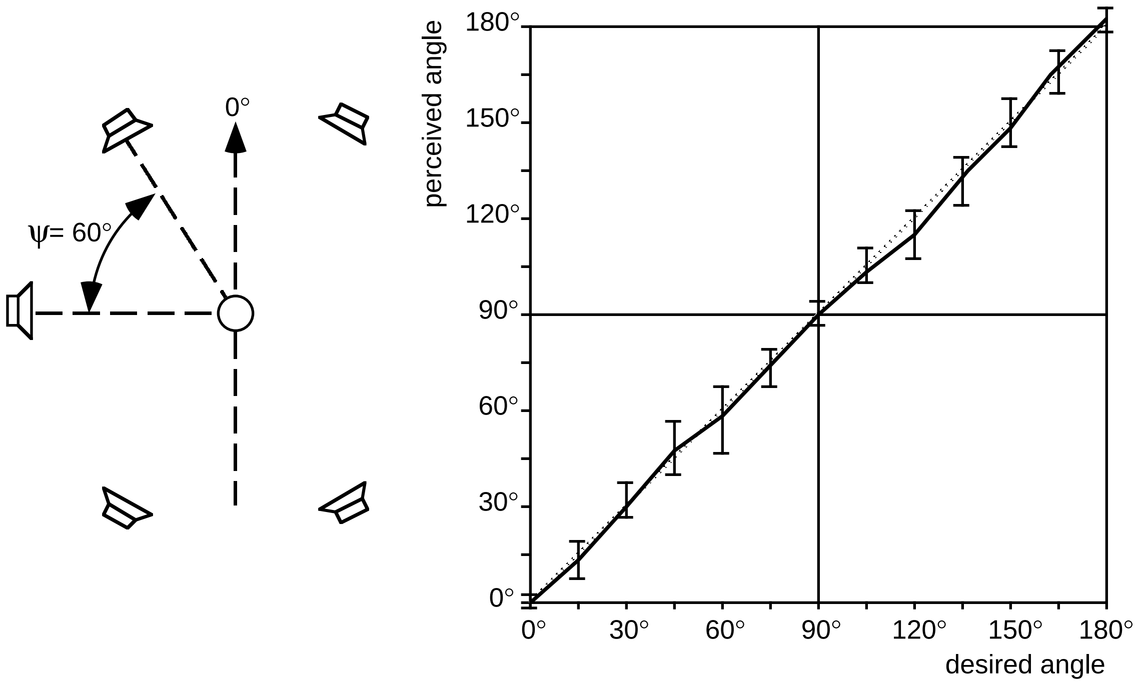


Figure 4.4 Perceived versus desired sound direction with noise signals, using 6 speakers arranged at 60 degree increments as shown on left [Theile-77]. The listener is facing forward.

These results clearly show that six loudspeakers would have been a better choice for the experimental setup. Nevertheless, experiments in early echo rendering were conducted using the four channel setup.

The first experiments consisted of rendering a rectangular room's early echo pattern. A 24 by 32 foot rectangular room was specified, along with a sound source location outside of the physical space, as shown below:

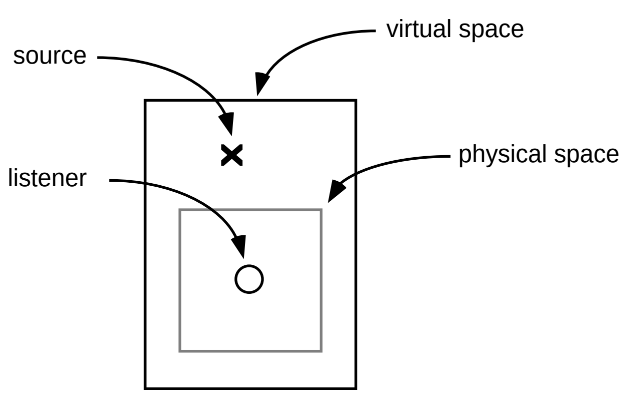


Figure 4.5 Rectangular virtual space and direct source location.

4.3 Optimizing the Early Echo FIR Filter

All virtual source locations were determined within a 200 foot radius, and these were used to create the FIR filter for each loudspeaker. The resulting FIR filters contained too many taps to be realized in realtime (40 taps was the maximum), thus pruning the FIR filters was necessary. This was done in two steps:

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

$$t_2 = \frac{t_0 a_0^2 + t_1 a_1^2}{a_0^2 + a_1^2} \tag{4.5}$$

$$a_2 = \sqrt{a_0^2 + a_1^2} \tag{4.6}$$

2) Filter taps of amplitude less than 0.01 (-20 dB) were deleted. If the resulting filters still contained too many taps, this step would be repeated with a higher threshold.

The resulting pruned filters contained roughly 20 taps per speaker. The pruning process had the effect of entirely eliminating distant virtual sources, as well as weak taps resulting from intensity panning. Thus, a virtual source that was angularly close to a speaker might be rendered entirely by that speaker after pruning.

4.4 Results of Early Echo Rendering

The early echo response of the room was auditioned by sending a monophonic channel of music to the input of the FIR filters. Note that the front left and right channels both contained a single tap corresponding to the direct sound source, and it was possible to switch from hearing only the direct monophonic source to hearing the source and early echoes, essentially turning the room on and off. The effect of enabling the room was quite pronounced. Adding the early echoes gave a real sense of the sound being enclosed within a space; the sound came from all around, the overall volume increased, the timbre became more resonant and hollow, and the spaciousness increased dramatically. Because the sound source was not on either symmetrical axis of the room, the early echo response was asymmetrical, thus providing uncorrelated lateral energy to the listener regardless of orientation.

Experimentation with larger virtual rooms revealed that the early echo pattern alone was a sufficient cue to distinguish among different sized rooms. However, the early echo response of the larger rooms suffered from an overly discrete sound to the echoes. The echo response to an impulsive sound was quite unrealistic, as if the sound was being reattacked, like a drum flam. This was clearly the result of oversimplification in 1) the room specification, 2) the modeling of reflections (equation 4.4), and 3) sound propagation through air.

Because the floor and ceiling reflections were not considered in the early echo rendering, the frequency response of the simulated room necessarily lacked features corresponding to floor to ceiling vibrational modes. It is unclear whether virtual sources created by floor and ceiling reflections need to be rendered by speakers that are overhead (or underneath) the horizontal plane of the listener. Rendering the out of plane virtual sources using the horizontal plane speakers would still contribute the desired features in the frequency domain, but the perceived direction of the virtual sources would be incorrect.

4.5 Modeling Air Absorption

One improvement to the 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 (at 50% humidity), the following equation was derived:

$$f_c = 2000^{\log_2 (d/75)} \tag{4.7}$$

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):

$$\bar{d} = c\frac{\sum_{i \in S} a_i^2 t_i}{\sum_{i \in S} a_i^2} \tag{4.8}$$

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 was used per FIR filter, based upon the mean echo distance of the entire FIR filter. Thus, there was 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. The problem of the multiple attacks was reduced, although it did not disappear entirely. Further improvements could be made by using more lowpass filters to better approximate the effects of air absorption, modeling the frequency dependent nature of wall reflections (which can be a significant phenomenon [Abbott-91]), and increasing the geometrical complexity of the virtual rooms, although these all require increased computational power to implement.

4.6 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.

4.7 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:

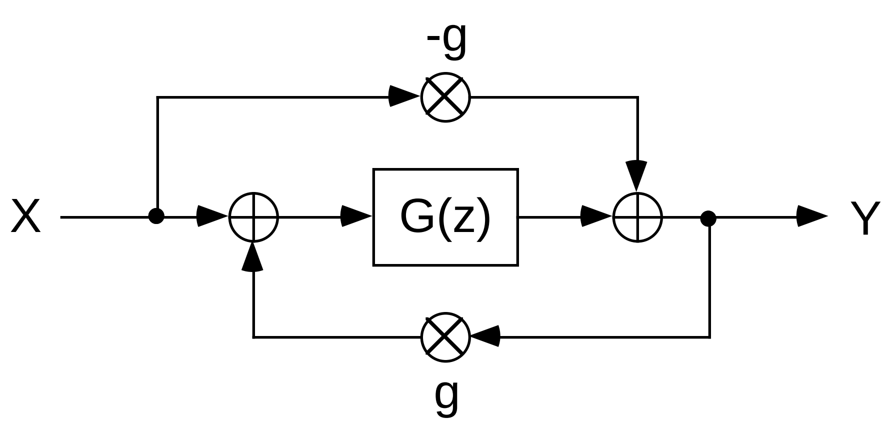


Figure 4.6 Allpass flow diagram. G(z) must be allpass.

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

$$H(z) = \frac{Y(z)}{X(z)} = \frac{G(z) - g}{1 - gG(z)} \tag{4.9}$$

The magnitude of H(z) is as follows:

$$|H(z)| = \sqrt{\frac{|G(z)|^2 - g\left(G(z) + G^*(z)\right) + g^2}{1 - g\left(G(z) + G^*(z)\right) + g^2|G(z)|^2}} \tag{4.10}$$

The magnitude of H(z) is unity if the magnitude of G(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 "pshhhh" (or a "bzzzz" with a different choice of delays and gains).

4.8 Nested Allpass Implementation

The allpass structure of figure 4.6 can be implemented easily by attaching operators to a sample delay line as shown below:

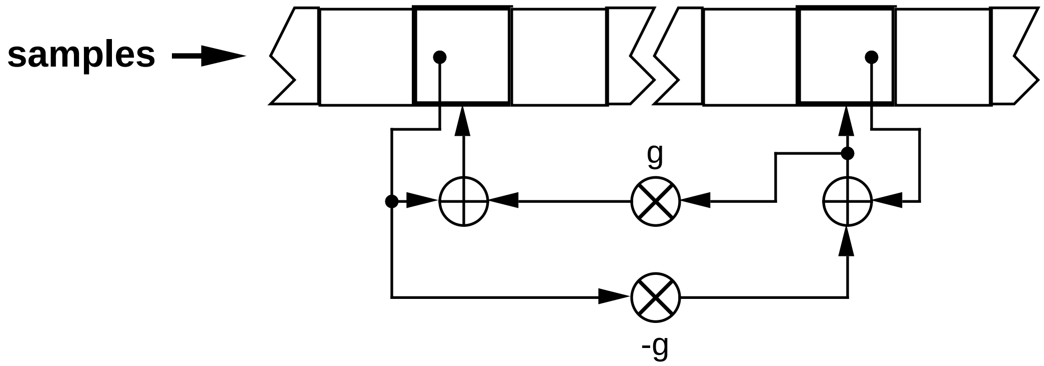


Figure 4.7 Allpass implementation using a sample delay line.

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:

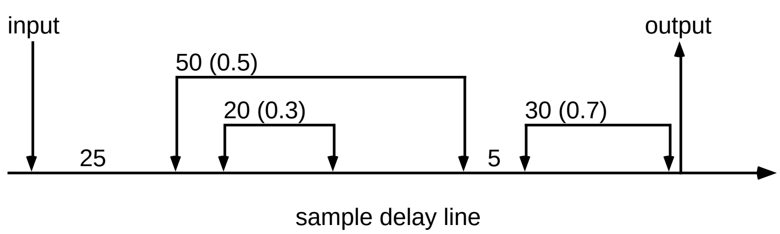


Figure 4.8 Example of schematic representation of an allpass reverberator.

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:

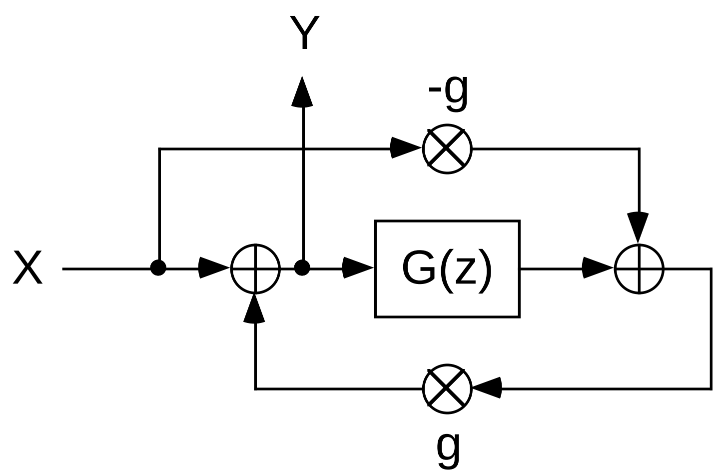


Figure 4.9 Flow diagram resulting from taking samples from interior of allpass delay line.

The z-transform of this system is:

$$H(z) = \frac{1-g^2}{1-gG(z)} \tag{4.11}$$

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 4.11, 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 4.11 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:

$$|H(z)| = \frac{1-g^2}{\sqrt{1-2g\,\mathrm{Re}\{G(z)\}+g^2|G(z)|^2}} \tag{4.12}$$

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:

$$|H(z)|_{\max} = 1+g \tag{4.13}$$

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.

4.9 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 simply 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:

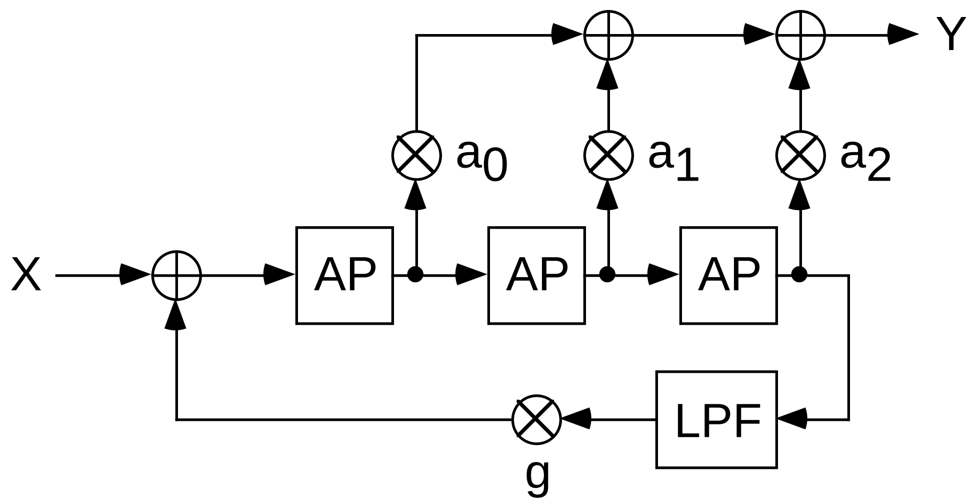


Figure 4.10 Generalized allpass reverberator with lowpass filtered feedback path and multiple weighted output taps.

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 4.7, 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.

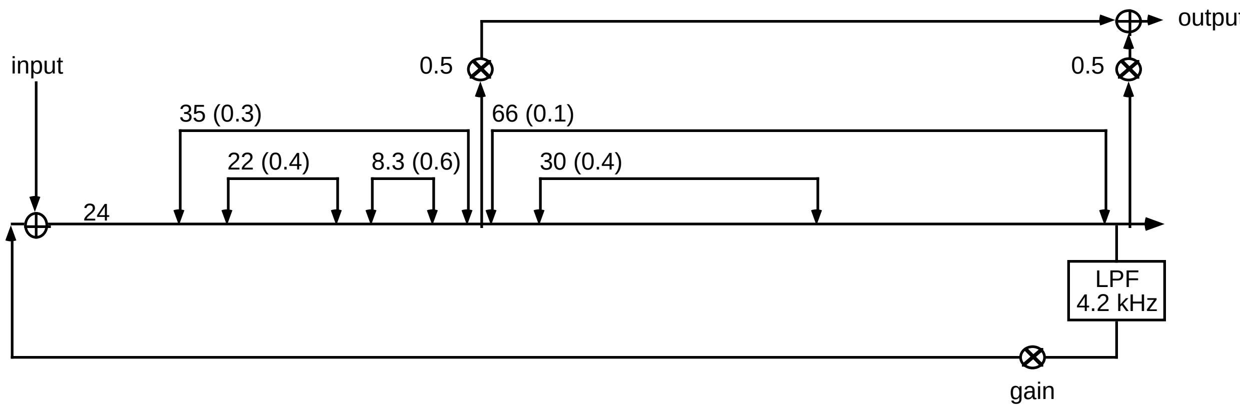
Obviously, there are a vast number of possible reverberators than can be built with the general structure of figure 4.10. If an automatic method could be devised to evaluate a reverberant response based on desired attributes, then reverberators could be designed using non-linear search techniques such as gradient descent, simulated annealing, and genetic algorithms. I experimented extensively with genetic algorithms to design reverberators, but the results of the searches were reverberators that scored well but sounded terrible. Clearly, the problem is to build an evaluator that hears reverberation the way a human does. I also used myself as an evaluator, but this was pointless, since the genetic search algorithms require thousands, if not millions, of evaluations to be performed. Nevertheless, I listened to many hundreds of different reverberator structures in this process. In the end, the design was done by trial and error.

4.10 Three Diffuse Reverberators

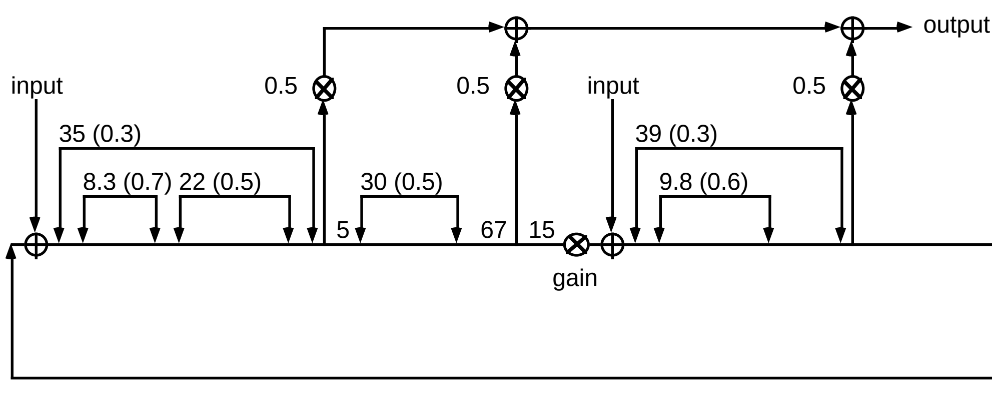
It was impossible to design a single diffuse reverberator to cover all desired reverberation times. A large room reverberator could not be made arbitrarily small by reducing the feedback gain; similarly, when a small room reverberator was given a large decay time by increasing g, it generally sounded bad. Thus, three different reverberators were designed to cover small, medium, and large rooms. The three reverberators are shown in figure 4.11. 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:

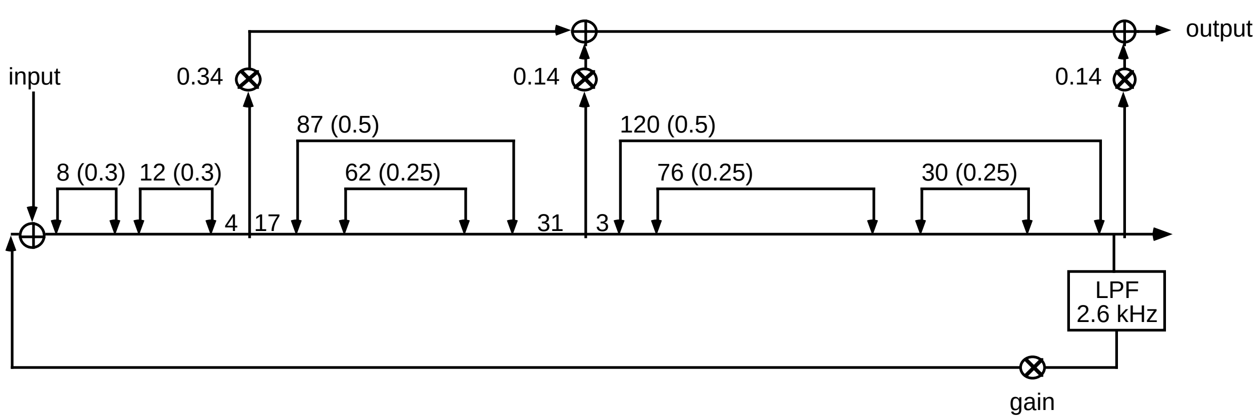


Figure 4.11 Diffuse reverberators for small, medium, and large rooms. See figure 4.8 for a description of these schematics.

4.11 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, four 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 four 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 could be disputed in terms of naturalness and timbre; in particular the small room reverberator sounded somewhat unnatural.

4.12 Combining Early Echoes with Diffuse Response

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

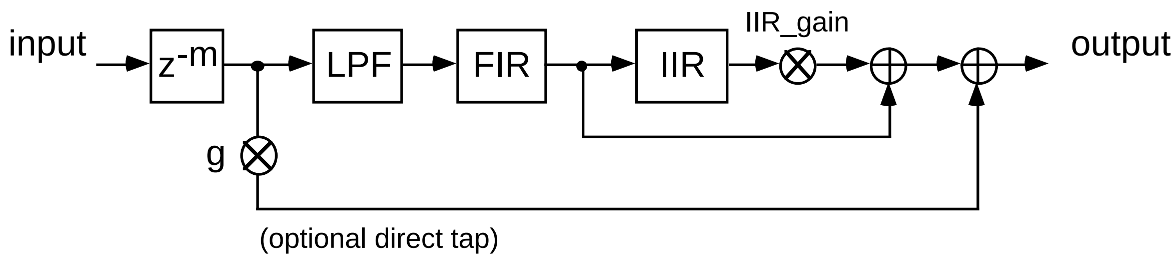


Figure 4.12 Combining FIR and IIR reverberators.

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.

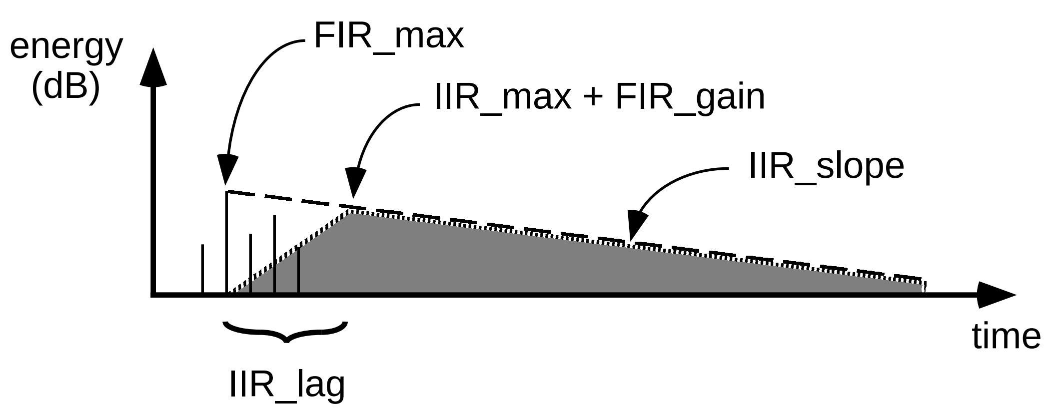


Figure 4.13 Combining FIR and IIR responses.

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:

$$\mathrm{IIR\_gain} = \mathrm{FIR\_max} + (\mathrm{IIR\_slope} \bullet \mathrm{IIR\_lag}) - (\mathrm{IIR\_max} + \mathrm{FIR\_gain}) \quad (4.14)$$

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 was 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.

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 were done by hand, the process is entirely deterministic and could be completely automated.

4.13 Results of Combined Listening

Four rooms were simulated: a 24 by 32 foot rectangular room with 10 foot ceiling, a 48 by 64 foot rectangular room with 15 foot ceiling, and two variations of an inverse fan shaped room, approximately 80 by 120 feet, with a 20 foot ceiling. Broadband wall reflection coefficients were set at 0.9 (which is somewhere between plaster and wood), and ceiling and floors were far more absorptive, typically 0.7 for floors (carpeting) and 0.8 for ceilings. For the inverse fan room, the two variations consisted of wall coefficients of 0.9 and 0.98, respectively, thus approximately simulating a change of wall material between plaster and concrete.

The calculated reverberation times of the three rooms were 0.69 seconds for the small room, 1.08 seconds for the medium room, 1.53 seconds for the large room with wooden walls, and 1.72 seconds with concrete walls. Note that the low ceiling and hence, small mean free path, limit the large room's reverberation time even with extremely reflective walls. The three basic room types used the small, medium and large room diffuse reverberators, respectively. The simulations were computationally restricted. Because the larger reverberators are somewhat complicated, only 6 early echoes could be rendered in the medium and large rooms, whereas 11 early echoes could be rendered in the small room.

Using a monophonic source of music, the four rooms were alternately auditioned. In order that the overall listening level remained constant, the source gain was made louder in the larger rooms. The three basic room types sounded completely different, and although each used a different diffuse reverberator, the change was just as prominent when only the early echoes were auditioned. The apparent size, timbre, and brightness of the rooms were largely determined by the early echo portion of the response.

Adding the diffuse reverberation rounded out the rooms, vastly increasing the spatial impression (especially overhead), and reducing the perception that the sound was coming from the four speakers. The diffuse reverberation was also very noticeable when the source material stopped abruptly, causing a pleasant and natural reverberant decay in the larger rooms. The difference between the concrete and normal large room was subtle, but predictable: the early echoes were more distinct and the room more reverberant with concrete walls. The small room was very interesting, if somewhat unnatural, sounding more like a tile bathroom than a living room. The medium room was not unpleasant, but was uninteresting.

The room impression was largely independent of orientation within the space, and it was possible to stray several feet from the center before noticing any difference in sound quality. As one moved further from the center, the sound of the closest speaker became dominant, ruining the spatial impression of the simulation.

It was possible (and enjoyable) to listen to the system for long periods of time without fatigue. In fact, after listening to music rendered through the simulated rooms, returning to ordinary stereo reproduction was a disappointment.

Some problems with the simulation were observed:

1) The setup was not in an acoustically neutral space. The acoustics of the Cube were somewhat noticeable during simulation.

2) The source material was not reverberation free. All of the recordings had natural or artificial reverberation applied during the recording/production process. It would have been preferable to use reverberation free material.

3) The source material was a monophonic channel from a stereo recording. This affected the naturalness of the listening experience.

In one experiment, the early echo pattern of a rectangular room was simulated with stereo inputs. Two separate source positions within the room were assigned to the left and right channel inputs. These sources spawned a set of left and right virtual sources, resulting in a left and right channel FIR filter per output speaker. The results of this test were excellent, but the computational demand prevented the addition of any diffuse reverberation.

4.14 The Reverb Compiler

A compiler was developed to convert reverberator specifications to efficient DSP code [Gardner-91]. The compiler, called *Reverb*, is a Macintosh application that provides the user with a multiple window text editor. Reverberator programs are entered in text format, compiled into DSP code, and downloaded into an Audiomedia DSP card for realtime execution. The user may switch freely between different Reverb programs. Programs may also be written to run on multiple DSP cards.

The Reverb programming language was specifically designed to implement reverberation algorithms, and is very simple and concise. Rather than specifying algorithms by combining functional blocks, Reverb programs are specified at a lower level by attaching operators directly to delay lines. The set of operators includes basic operations such as move, add, subtract, multiply, and more advanced operations such as FIR, comb, allpass, and lowpass filters. The compiler generates extremely efficient code, making full use of the features of the Motorola 56001 DSP.

An Audiomedia DSP card with expanded memory provides 24575 samples of delay, which is 557 milliseconds at the 44.1 kHz. sampling rate. Reverb programs can contain up to 20 comb filters, 14 allpass filters, or a 40 tap FIR filter (non-adjacent taps). These specifications were just adequate to implement the reverberators described in this chapter.

The Reverb program is available via anonymous FTP at the Internet address "cecelia.media.mit.edu". The program and a user's manual can be found in the subdirectory "reverb".

5. Conclusions and Future Work

I am quite pleased with the results of the room reverberation modeling. It is clear that convincing simulations of various sized rooms can be realized with a small set of loudspeakers surrounding the listener. More channels and computational power will make the simulations better. The next logical step is to increase the number of horizontal speakers to 6 (at 60 degree intervals). Also, the early echo rendering should include virtual sources created by floor and ceiling reflections. This will cause the frequency response of the simulated room to contain features that correspond to floor to ceiling vibrational modes. The addition of overhead speakers would permit a full three dimensional rendering of early echo patterns.

Another improvement to the room simulation would be to account for the frequency dependent nature of surface reflections. For most rooms, this phenomenon is more significant than air absorption, but it is harder to simulate efficiently. Perhaps a crude simulation can be rendered by one or more low order filters per output channel. The filter parameters would be derived by considering the angle dependent frequency response of all room surfaces in conjunction with the set of virtual sources.

The diffuse reverberation algorithms based on nested and cascaded allpass filters have proven to be quite useful, and they represent a significant improvement over the Schroeder style reverberator. Nevertheless, the study of reverberation algorithms, though fascinating, seems ultimately fruitless. This is because the advent of inexpensive, large, realtime convolution will render the IIR reverberator obsolete. In addition, the problems encountered have proven to be quite impenetrable. I was unable to develop any theory that related reverberator design principles to perception, beyond simple heuristics and common sense. I was also unable to develop an automatic reverberation evaluator, which would have enabled the use of non-linear search techniques to design reverberators. The latter problem deserves additional consideration, since developing hearing models is a worthwhile endeavor.

The use of FIR cancellation filters to cancel acoustic feedback clearly warrants further attention. The technique may not be appropriate for general sound reinforcement systems, but may still be useful for implementing a virtual acoustic room, whether operating alone or in conjunction with other techniques, such as time varying reverberation. This is because a virtual acoustic room can be designed to optimize the performance of the acoustic feedback cancellation system. This would involve controlling the following parameters:

- Acoustical properties of physical space. Desirable properties are a short reverberation time, linear room response, and low ambient noise. We would also like to minimize time variation in room response due to ventilation.

- Speaker and microphone placement. In addition to fulfilling the requirements of the reverberation rendering, we seek an arrangement that will minimize the variation in speaker to microphone response as people move about in the physical space.

- Architecturally imposed constraints on the motion of people within the space. Again, this is to minimize time variation in the room response, but can also be used to position people in ideal listening locations.

Without some method of eliminating acoustic feedback, an interactive virtual acoustic environment is impossible, particularly in regards to the proper rendering of early echo response. The research direction should be to develop means of measuring non-linearities and time variation in rooms, and to integrate noise, non-linearity, and time variation into our cancellation model. If the issues of non-linearity and time variation are successfully resolved, then the next step would be to first simulate, and then construct a functional virtual acoustic room.

An exciting possibility is the use of adaptive filters for acoustic feedback cancellation, especially if time variation in the room response is a significant problem for static cancellation systems. I suspect that the convergence time for the adaptive algorithms will be the limiting factor in the success of this technique. Since the convergence time improves when noiselike input signals are used [Sondhi-92], perhaps it will be possible to inject noise into the processed signal such that it speeds the convergence time of the adaptive algorithms, but is otherwise masked from audibility.

References

[Abbott-91]
James F. Abbott, “The Interaction of Sound and Shock Waves with Flexible Porous Materials,” Ph.D.. thesis, Department of Physics, MIT, Cambridge, MA (1991).

[Barron-81]
M. Barron and A. H. Marshall, "Spatial Impression Due to Early Lateral Reflection in Concert Halls: The Derivation of a Physical Measure," *Journal of Sound and Vibration*, 77 (2), (1981).

[Benade-85]
A. H. Benade, “From Instrument to Ear in a Room: Direct or via a Recording,” *J. Audio Engineering Society*, Vol 33, No. 4 (1985).

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

[Berkhout-88]
A. J. Berkhout, “A Holographic Approach to Acoustic Control,” *J. Audio Engineering Society,* Vol. 36, No. 12 (1988).

[Borish-84a]
Jeffrey Borish, "Electronic Simulation of Auditorium Acoustics," Ph.D.. thesis, Center for Computer Research in Music and Acoustics, Department of Music, Stanford University, CA, (1984).

[Borish-84b]
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-91]
William G. Gardner, “Reverb - A Reverberator Design Tool for Audiomedia,” unpublished users manual (1991). Available from Music and Cognition office at MIT Media Lab.

[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).

[Kuttruff-91]
Heinrich Kuttruff, *Room Acoustics*, Elsevier Science Publishing Company, New York, NY. (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).

[Moore-90]
Moore, F. Richard, *Elements of Computer Music*, Prentice-Hall, Englewood Cliffs, NJ (1990). Pages 353-359.

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

[Neely-79]
Stephen T. Neely and Jont B. Allen, "Invertibility of a room impulse response," *J. Acoustical Society of America*, 66 (1), (1979).

[Oppenheim-89]
Alan V. Oppenheim and Ronald W. Schafer, *Discrete Time Signal Processing*, Prentice Hall, Englewood Cliffs, NJ. (1989).

[Parkin-65]
P. H. Parkin and K. Morgan, "'Assisted Resonance' in the Royal Festival Hall, London,"*J. Sound Vib.* 2 (I), 74-85 (1965).

[Ratliff-74]
P. A. Ratliff, "Properties of Hearing Related to Quadraphonic Reproduction," BBC RD 38 (1974).

[Rife-87]
Douglas D. Rife and John Vanderkooy, "Transfer-Function Measurement using Maximum-Length Sequences," *J. Audio Engineering Society,* Preprint 2502 (1987).

[Sabine-00]
W. C. Sabine, "Reverberation," originally published in 1900. Reprinted in *Acoustics: Historical and Philosophical Development*, edited by R. B. Lindsay. Dowden, Hutchinson, and Ross, Stroudsburg, PA. (1972).

[Schafer-68]
Ronald W. Schafer, "Echo removal by discrete generalized linear filtering," Ph.D.. thesis, Electrical Engineering Department, MIT, Cambridge, MA (1968).

[Schroeder-54]
M. R. Schroeder, "Die Statistischen Parameter der Frequenzkurven von Grossen Raumen," *Acustica*, Vol 4, (1954). See [Schroeder-87] for English translation.

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

[Schroeder-63]
M. R. Schroeder and B. S. Atal, "Computer Simulation of Sound Transmission in Rooms," IEEE Int. Conv. Rec. 7, pp. 150-155 (1963).

[Schroeder-70]
M. R. Schroeder, "Digital Simulation of Sound Transmission in Reverberant Spaces," *J. Acoustical Society of America*, Vol. 47, No. 2 (1970).

[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).

[Schroeder-87]
M. R. Schroeder, "Statistical Parameters of the Frequency Response Curves of Large Rooms," *J. Audio Engineering Society*, Vol. 35, No. 5 (1987). English translation of [Schroeder-54].

[Sondhi-67]
Man Mohan Sondhi, "An Adaptive Echo Canceller," *The Bell System Technical Journal*, Volume XLVI, No. 3, (1967).

[Sondhi-91]
Man Mohan Sondhi, "Acoustic Echo Cancellation for Stereophonic Teleconferencing," IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, May (1991).

[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).

[Stockham-75]
Thomas G. Stockham, Jr., “Blind Deconvolution Through Digital Signal Processing,” Proceedings of the IEEE, Vol. 63, No. 4, (1975).

[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.