
**A hybrid model for simulating diffused first reflections
in two-dimensional acoustic environments**

Geoffrey Glen Martin

Faculty of Music, McGill University

March, 2001

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment
of the requirements of the degree of Doctor of Philosophy

© Geoffrey Martin, 2001

Abstract

Although it is widely accepted that the diffusion of early reflections in acoustic spaces intended for music performance greatly improves the perceived quality of sound, current manufacturers of synthetic reverberation engines continue to model reflecting surfaces as having almost perfectly specular characteristics. This dissertation describes a hybrid method of simulating diffusion based on both physical and phenomenological modeling components.

In 1979, Manfred Schroeder described a method of designing and constructing diffusing surfaces based on a rather simple mathematical algorithm which provides diffused reflections in predictable frequency bands. This structural device, now known as a “Schroeder diffuser,” has become a standard geometry used in constructing diffusive surfaces for spaces intended for music rehearsal, recording and performance. While it is possible to use DSP to model the characteristics of reflections off such a surface, a reflection model based exclusively on a surface constructed of a Schroeder diffuser has proven in informal tests to be as aesthetically inadequate as a perfectly specular model. Control of both the spatial and temporal envelopes of the diffusive reflection are required by an end user in order to tailor the reflection characteristics to the desired impression.

In 1974 an empirical model for computing light reflections off objects in a three-dimensional environment was developed by Phong Bui-Toung. This algorithm

incorporated both a specular and diffuse component with relationships controlled by an end user.

This dissertation describes the adaptation and implementation of the Phong shading algorithm in conjunction with a physical model of components of the Schroeder diffuser for the modeling of diffuse reflections in synthetic acoustic environments. The inclusion of the Phong algorithm provides precise control over the balance between the spectral and diffusive components of the reflection. In addition, directivity functions for sound sources and receivers in the virtual space are described.

Analysis and evaluation of the model using mathematical and empirical methodologies are discussed and stereo and multichannel audio examples produced by the system are included.

Résumé

Bien qu'il soit généralement admis que lors de l'exécution d'œuvres musicales dans des salles de concert, la diffusion des premières réflexions acoustiques améliore la qualité du son perçu, les constructeurs de dispositifs de réverbération artificielle persistent à modéliser des surfaces réfléchissantes n'ayant que des caractéristiques presque uniquement spéculaires. La présente thèse décrit une nouvelle méthode de simulation de diffusion acoustique hybride en ce qu'elle utilise des éléments de modèles physiques et phénoménologiques.

En 1979, Manfred Schroeder a décrit une méthode pour créer et construire des surfaces diffusantes en utilisant un algorithme mathématique fournissant des réflexions diffuses à l'intérieur de bandes de fréquences prévisibles. Cette structure, appelée un «diffuseur de Schroeder», est aujourd'hui un élément géométrique standard utilisé dans la construction des surfaces diffusantes destinées aux salles de répétition, aux studios d'enregistrement, et aux salles de concert. Bien qu'il soit possible d'utiliser des techniques de traitement de signaux pour modéliser les caractéristiques réfléchissantes d'une telle surface, un modèle de réflexion basé uniquement sur un «diffuseur de Schroeder» s'est avéré tout autant insatisfaisant sur le plan esthétique qu'un modèle entièrement spéculaire. Le contrôle des enveloppes spatiales et temporelles des réflexions diffuses est requis pour ajuster le son selon l'impression souhaitée.

En 1974, un modèle empirique utilisé pour calculer la réflexion de la lumière par des objets dans un environnement en tri-dimensionnel a été développé par Phong Bui-Toung. Cet algorithme incorpore des composants spéculaires et diffusés ayant un rapport contrôlable.

Cette thèse décrit l'adaptation de l'algorithme Phong à la création de réflexions diffusées dans des environnements acoustiques synthétiques, en combinaison avec un modèle physique des composantes d'un «diffuseur de Schroeder». L'ajout de l'algorithme Phong permet un contrôle précis du rapport entre les composantes spectrales et diffusantes des réflexions. En outre, des fonctions de directivité pour les sources sonores et les récepteurs situés dans l'espace virtuel sont décrites.

L'analyse et l'évaluation du modèle à l'aide de méthodes mathématiques et empiriques sont discutées, et des exemples sonores produits par le système sont inclus.

Dedication

This thesis is dedicated to my mother, Janet;
to my sisters, Andrée and Nicholle;
and to the memory of my father, Stewart.

Table of contents

Abstract	i
Résumé	iii
Dedication	v
Table of contents	vi
Acknowledgements and thanks	xiii
List of symbols and abbreviations	xiv
1 Introduction	1.1
1.1 Characteristics of sound and wave propagation	1.7
1.2 Elements of architectural acoustics	1.13
1.2.1 Reflection characteristics	1.13
1.2.1.1 Specular reflections	1.16
1.2.1.2 Diffused reflections	1.18
1.2.2 Surface characteristics	1.22
1.2.2.1 Irregular surfaces	1.23
1.2.2.2 Maximum length sequence diffusers	1.23
1.2.2.3 Schroeder diffusers	1.26
1.3 Perceptual significance of acoustic characteristics	1.33
1.3.1 Quantitative percepts	1.34

Table of contents

1.3.2 Qualitative percepts	1.38
1.4 Significance of acoustics simulation	1.41
1.5 Historical review of synthetic digital reverberation algorithms	1.44
1.5.1 Tapped recirculating delay model	1.45
1.5.2 Geometric model	1.47
1.5.2.1 Ray-trace model	1.49
1.5.2.2 Image model	1.50
1.5.2.3 Cone trace model	1.53
1.5.2.4 Hybrid method	1.55
1.5.3 Physical models	1.55
1.5.3.1 Boundary element method	1.57
1.5.3.2 Finite element method	1.57
1.5.3.3 Digital waveguide mesh	1.58
1.5.4 Convolution through a measurement of real space	1.61
1.5.5 Stochastic model	1.63
1.6 Review of documented methods of diffusing early reflections	1.65
1.6.1 Ray tracing with randomized ray directions	1.66
1.6.2 Ray tracing where the diffused energy decays with the reverberation time of the room	1.67
1.6.3 Ray / cone tracing where the diffused part of the energy initiates a D-D reflection process	1.67
1.6.4 Image source clouds	1.69

Table of contents

1.6.5 Filters on specular paths	1.70
1.6.6 Hybrid method combined with separate ray tracing	1.71
1.6.7 Early part, hybrid method / late part, modified ray tracing	1.71
1.6.8 Wall subdivision using specular and diffusing nodes	1.72
1.6.9 Early part image model plus S-D combinations, late part ray tracing	1.72
1.6.10 Split-up of rays reflected on diffusers	1.73
1.6.11 Digital waveguide mesh with modulating ringguides	1.75
1.6.12 Convolution through measured impulse responses	1.76
1.7 Typical implementation of early reflection patterns	1.78
1.7.1 Impulse response analyses	1.81
1.7.1.1 t.c. Electronic M3000.....	1.81
1.7.1.2 Lexicon 480L.....	1.87
1.7.2 General assumptions	1.95
2 Three-dimensional graphics models	2.1
2.1 Models of reflection	2.2
2.2 Shading algorithms	2.2
2.2.1 Gourad shading model	2.3
2.2.2 Phong shading model	2.4
3 Two-dimensional acoustic model	3.1
3.1 Phong shading model adaptation	3.2
3.1.1 Specular component	3.3

Table of contents

3.1.2 Diffused component	3.6
3.2 Instrument directivity	3.11
3.2.1 Instrument radiation patterns	3.11
3.2.2 Instrument rotation	3.13
3.3 Microphone sensitivity	3.19
3.3.1 Microphone polar patterns	3.19
3.3.2 Microphone rotation	3.21
3.4 Modeling wall details	3.23
3.4.1 Schroeder diffuser wells	3.23
3.5 Building the full impulse response	3.30
3.5.1 Specular reflection component	3.30
3.5.2 Diffused reflection component	3.31
3.5.3 Combining the specular and diffused components	3.37
3.6 Advantages and disadvantages of the system	3.39
4 Implementation and analysis	4.1
4.1 Gain error analysis	4.3
4.2 Propagation error and phase response	4.4
4.3 Creation of sound files for measurement and listening tests	4.7
4.4 Frequency response analysis.....	4.12
4.4.1 Entire impulse response	4.12
4.4.2 Waterfall plots	4.16

Table of contents

4.4.3 Spectrograms	4.20
4.5 Comparative cross-correlation measurements	4.23
4.5.1 Measurement and reproduction equipment	4.23
4.5.2 Measurement environment characteristics	4.26
4.5.3 Results	4.27
4.6 Imaging characteristics	4.30
4.6.1 Microphone implementation	4.32
4.6.2 Pair-wise power panning emulation	4.32
4.6.3 Ambisonics emulation	4.35
4.6.4 Binaural simulation / Auralization	4.37
5 Listening tests	5.1
5.1 Technical descriptions	5.3
5.1.1 Hardware and software	5.3
5.1.2 Sound sources	5.6
5.1.3 Equipment configuration and alignment	5.7
5.1.4 Test subjects	5.8
5.2 Test 1: A / B / X test	5.10
5.2.1 Results	5.12
5.3 Test 2: Mix preference test	5.16
5.3.1 Results	5.18
5.4 Analysis	5.23

6 Conclusion	6.1
6.1 Future developments	6.2
6.1.1 Aesthetic improvements	6.2
6.1.2 Ergonomic improvements	6.3
6.1.3 Technical considerations	6.7
6.2 Conclusion	6.10
Appendix A: Effects processor parameters	I
A.1 t.c. electronic M3000	I
A.2 Lexicon 480L	III
Appendix B: Technical details of sound source recordings	IV
Appendix C: MARLab acoustical characteristics	VI
Appendix D: Monitoring system response characteristics	XI
Appendix E: Listening test instructions	XXV
E.1 Test 1	XXV
E.2 Test 2	XXVI
Appendix F: Information regarding included CD-ROM	XXVII
F.1 Stereo demonstration	XXVII
F.1.1 Minimum system requirements	XXVIII
F.1.2 Instructions and configuration	XXVIII

Table of contents

F.2 Multichannel demonstration	XXIX
F.2.1 Minimum system requirements	XXIX
F.2.2 Instructions and configuration	XXIX
References and bibliography	XXXII

Acknowledgements and thanks

Thanks to the following individuals and companies:

Jason Corey and **René Quesnel**, for attempting to keep me sane in spite of living a hermetically-sealed existence in the basement of Redpath Hall. My advisors, **Dr. Philippe Depalle** and **Dr. Wieslaw Woszczyk**, for their guidance and support. **Dr. Mark Ballora**, for asking the right questions. **Dr. Søren Bech, Poul Praestgaard** and **Bang & Olufsen A/S**, for continued advice and technical support as well as the loan of audio playback equipment used daily in the MARLab. **Kim Rishøj, Morten Lave, Thomas Lund** and **t.c. electronic A/S**, for the loan of invaluable equipment to the MARLab and continual technical support. **Dr. Takeo Yamamoto** and **Pioneer Corporation**, for financial support to the MARLab. Students of the Graduate Program in Sound Recording at McGill who participated in the listening tests. **Jane Giffen**, for taking the time to carefully proofread this document, checking for both grammatical errors and split infinitives. **Jennifer Stephenson**, for putting up with me while I wrote this and ensuring that I remembered to eat at least one meal every day.

List of symbols and abbreviations

\wp	instantaneous pressure
\wp_0	equilibrium pressure
α	absorption coefficient
δ	diffusion coefficient
γ	angle to receiver from centre of cone in cone trace
γ_0	angular spread of cone in cone trace
ζ_i	instrument rotation
ζ_m	microphone rotation
ϑ_i	angle of incidence
ϑ_r	angle of reflection
λ	wavelength
λ_0	design wavelength of diffuser
ξ_a	phantom image angle of sound source “a”
ρ_0	volume density
σ_i	angle of radiation of instrument
σ_m	angle of incidence to microphone

List of symbols and abbreviations

ϕ	phase shift
χ^2	chi-square test for independence
ψ_p	angle between the viewer and the angle of reflection
Ψ_b	angular location of loudspeaker “b” in Ambisonics system
ω	angular frequency
a	sound source in Ambisonics system
b	loudspeaker in Ambisonics system
B	total number of loudspeakers in Ambisonics system
B_{min}	minimum number of loudspeakers for Ambisonics system
c	speed of sound
<i>diffuse</i>	power level of diffused reflection component
d_n	pipe length or well depth in a MLS or Schroeder diffuser
D	distance
D_s	reflection delay time
D_{tot}	total distance travelled by reflection
E_d	delay error
f	frequency
F_s	Sampling rate
g	instrument directivity coefficient
G	gain

List of symbols and abbreviations

G_d	gain of diffused reflection
G_s	gain of specular reflection
h	integer
H	number of cones in a cone tracing sphere
I	sound intensity
I_a	ambient intensity
I_i	incident intensity
I_r	reflected intensity
\overline{IR}	distance between the instrument and the point of reflection
I_{tot}	total intensity
k	wave number
k_a	ambient radiance scalar
k_d	diffused reflection radiance scalar
k_s	specular reflection radiance scalar
m	order of pantaphonic Ambisonics system
n	well number in a Schroeder diffuser
N	index of modulus in a Schroeder diffuser
p	instantaneous sound pressure
P	maximum peak sound pressure
P_a	pressure of sound source in Ambisonics system
P_e	effective sound pressure
P_m	pressure component of microphone

List of symbols and abbreviations

r	acoustic resistance
R	pressure reflection coefficient
\overline{RM}	distance between the point of reflection and the microphone
r_n	radius of a pipe
s	index of smoothness
s_n	sequence of relative well depths in Schroeder diffuser
S_n	cross sectional area of a pipe
t	time in the impulse response measured in samples
T	pressure transmission coefficient
T_s	sampling period
u	instantaneous particle velocity
U	maximum peak particle velocity
U_a	second-order Ambisonics B-format channel
V_a	second-order Ambisonics B-format channel
w	well width in a Schroeder diffuser
W_a	zeroth-order Ambisonics B-format channel
x	acoustic reactance
X_a	first-order Ambisonics B-format channel
X_I	X-axis value in Cartesian location of instrument
X_M	X-axis value in Cartesian location of microphone
X_R	X-axis value in Cartesian location of point of reflection
Y_a	first-order Ambisonics B-format channel

List of symbols and abbreviations

Y_I	Y-axis value in Cartesian location of instrument
Y_M	Y-axis value in Cartesian location of microphone
Y_R	Y-axis value in Cartesian location of point of reflection
z	acoustic impedance
z_d	acoustic impedance of the cap closing a pipe
z_o	specific acoustic impedance
Z_I	Z-axis value in Cartesian location of instrument
Z_M	Z-axis value in Cartesian location of microphone
Z_n	acoustic impedance at the mouth of a closed pipe
Z_R	Z-axis value in Cartesian location of point of reflection
Z_o	acoustic impedance of a plane wave in a pipe

AP	Audio Precision
AQI	Acoustic quality index
B&K	Brüel & Kjær
B&O	Bang & Olufsen A/S
CAD	Computer assisted design
DDI	DDL Directivity Interface
DFT	Discrete Fourier transform
DSP	Digital signal processing
FFT	Fast Fourier transform

List of symbols and abbreviations

FIR	Finite impulse response
HRTF	Head-related transfer function
HTS	Head and torso simulator
IACC	Interaural cross correlation
IFFT	Inverse fast Fourier transform
IIR	Infinite impulse response
ITU	International Telecommunication Union
MARLab	Multichannel Audio Research Laboratory
MLS	Maximum length sequence
MLSSA	DRA Laboratory's Maximum length sequence system analyzer
MotU	Mark of the Unicorn
PSD	Power spectral density
QRD	Quadratic residue diffuser
RGB	Red, Green, Blue
RT_{60}	reverberation time
SDI	Surface diffusivity index
SPL	Sound pressure level
TRD	Tapped recirculating delay

1 Introduction

Science, which has become a great power in the last century, has analyzed everything divine handed down to us in the holy books. After this cruel analysis the learned of this world have nothing left of all that was sacred. But they have only analyzed the parts and overlooked the whole, and indeed their blindness is marvelous.

Fyodor Dostoevsky – The Brothers Karamazov

Soun is noght but air y-broken

Geoffrey Chaucer – The House of Fame

For the past 4 years, a team comprising the Multichannel Audio Research Laboratory (MARLab) at McGill University has been developing a real-time dynamic multichannel acoustics synthesizer dubbed “SceneBuilder” (Quesnel 1999). The mandate of this initiative is the development of a multichannel reverberator with controllers based on perceptual rather than physical attributes of a simulated acoustic environment. Initially comprised of an assembly of off-the-shelf digital mixers and reverb processors controlled using MIDI by a central proprietary software package, the system has grown to include custom digital signal processing (DSP) modules as required. The initial goal of the

MARLab was to use SceneBuilder to create virtual acoustic environments called “Scenes” using various parameter configurations based almost exclusively on aesthetic decisions, thus constructing *empirical models* (Edwards 1989) of various rooms and enclosures. Using relatively simple processes such as delay and equalization in addition to 8 parallel uncorrelated stereo reverberation streams applied to a monophonic anechoic recording, a number of environments were created by combining multiple components of each Scene. These individual components typically corresponded to locations in a concert hall such as the stage and audience areas, although they grew to include more nebulous perceptual components such as the “texture” of sound sources.

Although these efforts met with some success, the task proved decidedly monumental. In response to this, the system was modified to Version 2 which included a number of *physical model* components (Edwards 1989), providing the user with a group of parameter settings calculated from known physical laws and principles based on chosen physical attributes. These calculated parameters could then be used as a base set which could be subsequently modified according to aesthetic considerations. The topology of this system permitted the user to modify attributes such as sound source position as well as the location and absorptive characteristics of walls. These variables were used in the automatic calculation and update of various parameters such as the length and spectral balance of reverberation. In addition, the spatial positions of the sound source and individual first- and second-order reflections were correctly located around the listening position using a custom-developed panning algorithm (Martin 1999a). Appropriate delay

times and resulting Doppler shifts were implemented using multiple interpolated delay lines. These settings were subsequently fine-tuned using proprietary perceptual controllers designed for the system to control simultaneously multiple parameters, as well as to adjust them individually.

This version of SceneBuilder provided an excellent tool for the early development of a number of controls based on fundamental perceptual attributes such as spaciousness and depth. In addition, it proved extremely useful in improving on the characteristics of existing commercial systems such as sound source localization in the virtual space. It was found, however, that despite the system's superiority to other synthetic reverberation engines, comparisons between a virtual acoustical environment and its real-world counterparts highlighted many flaws. In one example, a multichannel recording of a marimba in McGill's Pollack Concert Hall was directly compared with a synthetic "equivalent" with humbling results. The simultaneous impressions of space, coherence and reverberation density from the real recording were unmatched by the SceneBuilder version. This experiment led to the third and current version of the system.

In Version 3, the topology is based on an arrangement similar to that which was used to produce the recording in Pollack Hall. It was reasoned that, in order to match the desirable qualities of the real recording, similar procedures should be employed. This required a revision of the basic implementation of the system. The mandate shifted from the transformation of the listening space into a virtual environment, to the placement of virtual microphones in the synthetic acoustic space with feeds to the real monitoring

system. These virtual microphone modules are controlled with various parameters that permit the user not only to locate them anywhere within the boundaries of the virtual space, but to alter characteristics such as polar pattern and rotation.

In all versions of the SceneBuilder system developed to date, early reflection patterns, whether positioned using empirical or deterministic models, have been produced using simple digital delay lines; either multiple individual delays in parallel or single multitap modules, chosen according to availability and efficiency. This arrangement, although an improvement on other digital reverberation systems due to the use of appropriate locations and gains, remains a major issue to be resolved in the comparison tests with recordings of instruments in real enclosures particularly due to the unnatural, perfectly specular reflection model produced by a simple delay line. The density and spread of diffused reflections contributes greatly to the beauty both of the apparent direct sound and the characteristics of the surrounding environment. As will be discussed, this is a concept which has long been accepted to be the case in real environments, but which continues to be largely ignored by almost all manufacturers of reverberation units, albeit with partially acceptable alibis.

Some initial efforts were made in SceneBuilder V.2 to create an impression of diffuse early energy using various configurations of first-order allpass filters applied to the early reflections. In this implementation, the early reflection set was summed to an array of eight intermediate mixing busses – one for each output channel. Each of these busses was processed using a first-order allpass filter with a fixed delay and variable gain. The

individual feedback gains of the filters were distributed among various positive and negative multiples of a scalar determined by a single “spaciousness” fader on the user interface. These preliminary experiments met with some success in that the method proved to be of use in providing some sense of increased width in the reflection set. This is due in large part to negative correlation between the multiple mixing busses. It was found, however, that the use of first-order allpass filters resulted in a number of audible problems largely due to excessive simplicity. As has been noted in other research, the flat frequency response of a filter regrettably does not necessarily imply its perceptual transparency (Moorer 1979). While the allpass arrangement makes no spectral changes, the transient response is highly altered. This effect was, in fact, an intentional modification, an attempt to “smear” the time response of the reflected energy. Unfortunately however, the model proved to be inadequate.

It became obvious through the various implementations of the SceneBuilder platform that a cluster of appropriately panned delays is required to simulate the early reflections of the synthetic environment. This discovery has since been independently confirmed (Lund 2000). It was also evident, however, that a simple multitap delay with low-complexity filtering, regardless of the relationship of the individual gains and times of the delay taps, was inadequate for the construction of aesthetically acceptable simulations. Investigations documented below in Section 1.7 also led to the realization that almost all commercial reverberation engines suffer from all of the same problems as SceneBuilder in this area. As a result, the development of a system that could simulate diffusive characteristics of

reflective surfaces was required. Some recent work in the simulation of diffused reflections off various specific surface geometries has been accomplished (Angus 1999) using a completely physical model for the analysis of real acoustical situations. Although systems discussed below have been developed for use with non-real-time auralization and predictive acoustics software packages, very few provide a realistic model of low-order diffused reflections and even fewer have been developed for reverberation engines. What is required is a method for modifying the diffusive characteristics of early reflections in a synthetic environment for music production – one that allows a user to manipulate various physical and perceptual characteristics of the process according to the requirements of the production.

The goal of this thesis is to develop a feasible procedure based on a hybrid of empirical, and physical models for the creation and manipulation of diffuse characteristics of first-order reflections in a virtual acoustic environment. Although, as is the case in many procedures used in the creation of recorded music, this goal appears at the outset to be very much an analysis of “the parts” – or possibly even just a single part – its contribution to the whole has proved in practice to be significant.

1.1 Characteristics of sound and wave propagation

The presence of sound in a transmission medium is, at its most basic level, a change in pressure over time. Any given medium has a normal instantaneous stasis pressure level \wp_o which is related to the mean distance between adjacent particles in the substance. In the case of air, this pressure is commonly known as the *atmospheric* or *barometric pressure* and is typically on the order of 100 kPa at sea level. When all of the particles in a gaseous medium in a given volume are at normal pressure, then the material is at its *volume density* or *constant equilibrium density*, abbreviated ρ_o and measured in kg/m³.

The instantaneous amplitude of sound level p is a measurement of the deviation of the instantaneous pressure \wp from this point of equilibrium \wp_o and is calculated using Equation 1.1 (Kinsler 1982).

$$p = \wp - \wp_o \quad (1.1)$$

A sinusoidal oscillation of this pressure reaches a maximum peak pressure P which determines the *sound pressure level* or *spl*. In air, this level is typically expressed as a logarithmic ratio of the effective pressure P_e referenced to the *threshold of hearing*, the commonly-accepted lowest sound pressure level audible by humans at 1 kHz, 20 microPascals, using Equation 1.2 (Woram 1989).

$$SPL = 20 \log_{10} \left[\frac{P_e}{20 \times 10^{-6} Pa} \right] \quad (1.2)$$

where, in the case of sinusoidal waveforms

$$P_e = \frac{P}{\sqrt{2}} \quad (1.3)$$

When a tuning fork is struck it vibrates with a regular periodic, or repetitive motion. The tines of the fork, moving in opposite directions, bend inwards and outwards, causing compressions and rarefactions due to the motion of the adjacent air particles. The frequency of these pressure changes f is measured in cycles per second, or “Hertz” where 1 Hz corresponds to 1 cycle per second. If there are no harmonics above the fundamental frequency present in the sound, then the change in air pressure over time follows a simple sinusoidal pattern and the sound source is said to be a *simple harmonic oscillator*. The addition of multiple harmonics above this fundamental frequency produces a resulting change in the waveform. By corollary, any waveform can be deconstructed into its constituent sinusoidal harmonic (or enharmonic) components at various frequencies, amplitudes and relative phases.

The characteristics of the waveform and the relationship between its harmonics can be expressed in a number of different ways. Possibly the most commonly-seen display is a Cartesian plot of the change in the instantaneous pressure p over time. In order to create a change in the pressure, a displacement of the particles from their stasis location is required. This, in turn, implies an instantaneous particle velocity u which, in linear acoustics, is independent of the wavefront velocity. A sinusoidal oscillation results in a peak velocity U which is proportional both to the peak amplitude of the pressure change

1.1 Characteristics of sound and wave propagation

and to the frequency in unbounded air. Since the particle velocity is dependent on the change in displacement, it is equivalent to the instantaneous slope, or the partial derivative of the displacement function. As a result, the velocity wave precedes the displacement wave by $\pi/2$ radians as is shown in Figure 1.1.

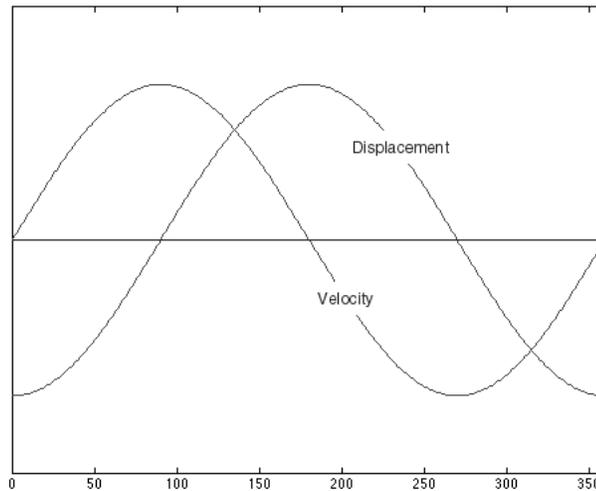


Figure 1.1 Relationship between particle displacement and velocity for a sinusoidal waveform

The relationship between the instantaneous particle pressure and velocity is determined by the *acoustic impedance* of the transmission medium. This is a ratio of the pressure to the particle velocity in unbounded air as is shown in Equation 1.4 (Kinsler 1962).

$$z = \frac{p}{u} \tag{1.4}$$

where z is the acoustic impedance in acoustic ohms.

In the specific case of unbounded plane waves, this ratio is also equal to the product of the volume density of the medium, ρ_o and the speed of wave propagation c as is shown in Equation 1.5 (Olson 1957). This value z_o is known as the *specific acoustic impedance* or *characteristic impedance* of the medium.

$$z_o = \rho_o c \quad (1.5)$$

where ρ_o is the volume density in kg/m^3 and c is the speed of wavefront propagation (also known as the speed of sound) in m/s .

In the rare case of progressive plane waves, the specific acoustic impedance of a medium is a real quantity. However, in the case of diverging waves, or more importantly for specific components in this dissertation, standing plane waves, it typically includes a complex component due to incurred shifts in phase. In this case, z is comprised of the sum of the specific acoustic resistance r and the specific acoustic reactance x as is expressed in Equation 1.6 (Kinsler 1982).

$$z = r + jx \quad (1.6)$$

In the case of a single point source, there is a finite amount of power distributed over an expanding spherical wavefront. As a result, as the wavefront propagates, that fixed amount of power is distributed over an increasing surface area. This area expands as a function of the square of the radius, and therefore the *intensity* of the sound – a measure of the sound power distributed over a fixed area perpendicular to the direction of

propagation – is inversely proportional to the square of the propagation distance of the wavefront. Since the power is proportional to the square of the amplitude, the sound pressure level is inversely proportional to the propagation distance. The standard distance used in measuring sound pressure levels is 1 m, therefore, for the purposes of this dissertation, we can calculate the attenuation due to propagation of an unbounded spherical wavefront using Equation 1.7.

$$G = \frac{1}{D} \tag{1.7}$$

where G is the gain applied to the signal and D is the distance of propagation from the sound source in metres.

In practice, for lower frequencies, no energy will be lost in the propagation through air. However, for shorter wavelengths, there is an increasing attenuation due to viscothermal losses in the medium. These losses in air are on the order of 0.1 dB per metre at 10 kHz as is shown in Figure 1.2.

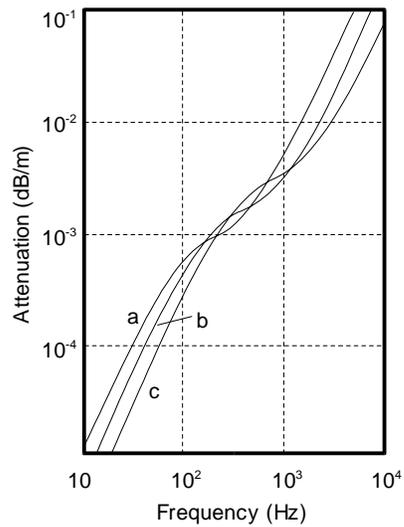


Figure 1.2 Attenuation resulting from air absorption due to propagation distance vs. frequency for various relative humidity levels (a) 20 %; (b) 40 %; (c) 80 % (Kuttruff 1991)

Since these losses are low for a relatively small propagation distance, they will be ignored for the direct sound and first reflections in this dissertation.

1.2 Elements of architectural acoustics

The properties of sound waves are most easily considered in the very specific case of a plane wave in a free field. This situation almost never exists in the real world – we are surrounded by wavefronts of an infinite number of shapes and propagation directions due in large part to reflections off the surfaces of surrounding objects. The study of architectural acoustics is concerned with the behavior of sound in such enclosed spaces.

This investigation can be carried out in a number of different interrelated domains. On a most basic perceptual level, we typically consider acoustic events in the *spatial domain*. That is to say, we are initially concerned with the angular direction and distance to a given sound source or reflection. Typically, the physical analysis of sound takes place in the interrelated *temporal, frequency and phase domains*. The specific characteristics of a pressure wave in the time domain determine its constituent frequency components and phase relationships and vice versa.

1.2.1 Reflection characteristics

When the wavefront of a propagating sound encounters a change in the propagation medium and the direction of wave propagation is perpendicular to the boundary of the two media, the pressure in the wave is divided between two resulting wavefronts – the transmitted sound, which passes into the new medium, and the reflected sound, which is transmitted back into its medium of origin as is shown in Figure 1.3 and expressed in Equation 1.8 (Kinsler 1982).

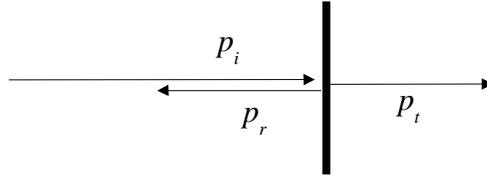


Figure 1.3 The relationship between the incident, transmitted and reflected pressure waves assuming that all rays are perpendicular to the boundary.

$$p_i + p_r = p_t \quad (1.8)$$

where p_i is the incident pressure in the first medium, p_r is the reflected pressure in the first medium and p_t is the pressure transmitted into the second medium at the surface. Similarly, the difference between the incident and reflected particle velocities equals the transmitted particle velocity as is shown in Equation 1.9 (Kinsler 1982).

$$u_i - u_r = u_t \quad (1.9)$$

As a result, we can combine Equations 1.4, 1.8 and 1.9 to produce Equation 1.10 (Kinsler 1982).

$$\frac{p_i + p_r}{u_i - u_r} = \frac{p_t}{u_t} = z_t \quad (1.10)$$

where z_t is the acoustic impedance of the second medium.

The ratios of reflected and transmitted pressures to the incident pressure are frequently expressed as the *pressure reflection* and *pressure transmission coefficients*, R and T , shown in Equations 1.11 and 1.12 (Kinsler 1982).

1.2.1 Reflection characteristics

$$R = \frac{P_r}{P_i} \quad (1.11)$$

$$T = \frac{P_t}{P_i} \quad (1.12)$$

Assuming that the impedance of the first medium is comprised of only a real component, it can therefore be shown that (Kinsler 1982)

$$z_t = r_i \frac{1+R}{1-R} \quad (1.13)$$

and therefore (Kinsler 1982)

$$R = \frac{(r_t - r_i) + jx_t}{(r_t + r_i) + jx_t} \quad (1.14)$$

where r_i and r_t are the acoustic resistances of the first and second media respectively and x_t is the acoustic reactance of the second medium.

As can be seen from this equation, if the incident pressure is positive and $r_i < r_t$, the reflection coefficient pressure is positive and there is no polarity shift of the reflection relative to the incident pressure wave. If $r_i > r_t$, the reflected pressure is negative and a resulting polarity shift occurs. If $r_i = r_t$, $R = 0$ and there is therefore no reflection.

Equation 1.14 also indicates that there will be a phase change in the pressure wave reflected off a reactive surface. This reflected wave may either lead or lag the incident wave by an amount between 0 and π radians at the boundary of the two media.

1.2.1.1 Specular reflections

The discussion in the previous section assumes that the wave propagation is normal, or perpendicular, to the surface boundary. In most instances, however, the *angle of incidence* – an angle subtended by a normal to the boundary (a line perpendicular to the surface and intersecting the point of reflection) and the incident sound ray – is an oblique angle. If the reflective surface is large and flat relative to the wavelength of the reflected sound, there exists a simple relationship between the angle of incidence and the *angle of reflection*, subtended by the reflected ray of sound and the normal to the reflective surface. *Snell's law* describes this relationship as is shown in Figure 1.4 and Equation 1.15 (Isaacs 1990).

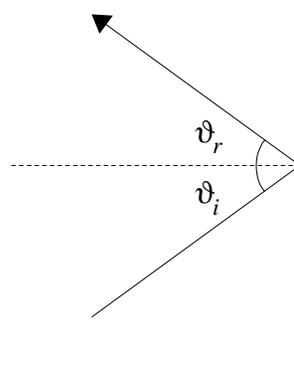


Figure 1.4 Relationship between the angles of incidence and reflection in the case of a specular reflector.

$$\sin(\vartheta_i) = \sin(\vartheta_r) \quad (1.15)$$

and therefore

1.2.1.1 Specular reflections

$$\vartheta_i = \vartheta_r \tag{1.16}$$

Since this type of reflection is most commonly investigated as it applies to visual media and thus reflected light, it is usually considered only in the spatial domain as is shown in the above diagram. The study of specular reflections in acoustic environments also requires that we consider the response in the time domain as well. This is not an issue in visual media since the speed of light is effectively infinite in human perception. If the surface is a perfect specular reflector with an infinite impedance, then the reflected pressure wave is an exact copy of the incident pressure wave. As a result, its impulse response is equivalent to a simple delay with an attenuation determined by the propagation distance of the reflection as is shown in Figure 1.5.

1.2.1.1 Specular reflections

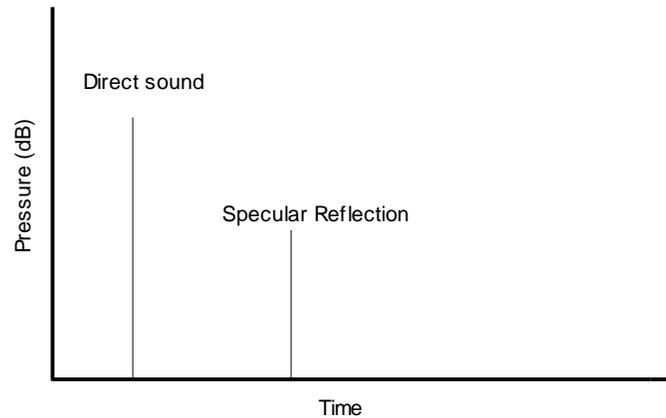


Figure 1.5 Impulse response of direct sound and specular reflection. Note that the Time is referenced to the moment when the impulse is emitted by the sound source, hence the delay in the time of arrival of the initial direct sound.

1.2.1.2 Diffused reflections

If the surface is irregular, then Snell's Law as stated above does not apply. Instead of acting as a perfect mirror, be it for light or sound, the surface scatters the incident pressure in multiple directions. If we use the example of a light bulb placed close to a white painted wall, the brightest point on the reflecting surface is independent of the location of the viewer. This is substantially different from the case of a specular reflector such as a polished mirror in which the brightest point, the location of the reflection of the light bulb, would move along the mirror's surface with movements of the viewer. *Lambert's Law* describes this relationship and states that, in a perfectly diffusing reflector, the intensity is proportional to the cosine of the angle of incidence as is shown in Figure 1.6 and Equation 1.17 (Isaacs 1990).

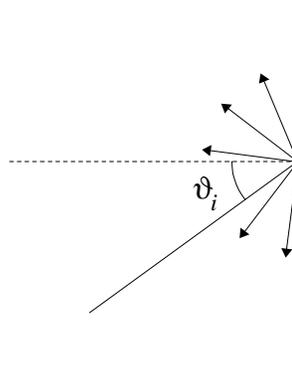


Figure 1.6 Relationship between the angles of incidence and reflection in the case of a diffusive reflector

$$I_r \propto I_i \cos(\vartheta_i) \quad (1.17)$$

where I_r and I_i are the intensities of the reflected and incident sound waves respectively.

In order to clarify the exact topic of this dissertation, a short discussion of some similar terms is required. Unfortunately, there is a lack of consistency in the terms commonly used to describe the diffusion of reflections (Dalenbäck 1994). The word “diffusion” is a term which is occasionally used in place of “diffused reflection.” However, diffusion is more correctly a result of multiple diffused reflections. Dalenbäck points out that the frequently-used term “scattering” has multiple meanings, one of which corresponds with diffused reflections. In general linear acoustics, “scattering” is frequently used to describe the result of diffraction whereas in applied acoustics its definition is generally understood to be a reflection from a rough surface. This second definition also includes an event known primarily to those studying optical reflections as “backscattering” in which the

1.2.1.2 Diffused reflections

wavefront is reflected back in the direction of the sound source at angles of incidence other than a normal to the reflecting surface.

For this dissertation, the term *diffused reflection* will refer to a reflection off an irregular surface.

There are a number of physical characteristics of diffused reflections that differ substantially from their specular counterparts. This is due to the fact that, whereas the received reflection from a specular reflector originates from a single point on the surface, the reflection from a diffusive reflector is distributed over a larger area as is shown in Figure 1.7.

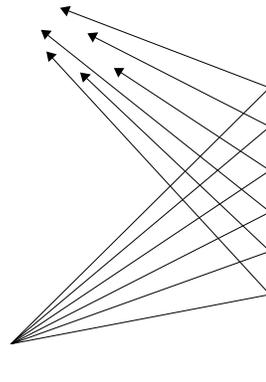


Figure 1.7 Diffused reflection showing spatial distribution of the reflected power for a single receiver.

Dalenbäck (1994) lists the results of this distribution in the spatial, temporal and frequency domains as the following:

- 1 *Non-specular regions are covered*
- 2 *temporal smearing and amplitude smoothing*
- 3 *reception angle smear*
- 4 *directivity smear*
- 5 *frequency content in reflection is affected*
- 6 *creation of a more uniform reverberant field*

The first issue will be discussed in Section 1.3.1. The second, third and fourth points are the product of the fact that the received reflection is distributed over the surface of the reflector. This results in multiple propagation distances for a “single” reflection as well as multiple angles and reflection locations. Since the reflection is distributed over both space and time at the listening position, there is an effect on the frequency content. Whereas, in the case of a perfect specular reflector, the frequency components of the resulting reflection form an identical copy of the original sound source, a diffusive reflector will modify those frequency characteristics according to the particular geometry of the surface. Finally, since the reflections are more widely distributed over the surfaces of the enclosure, the reverberant field approaches a perfectly diffuse field more rapidly.

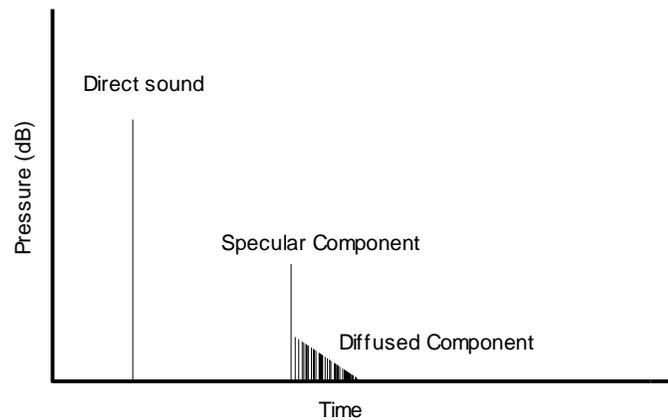


Figure 1.8 Impulse response of direct sound as well as the specular and simplified diffused reflection components. Note that the Time is referenced to the moment when the impulse is emitted by the sound source, hence the delay in the time of arrival of the initial direct sound.

1.2.2 Surface characteristics

The relative balance of the specular and diffused components of a reflection off a given surface are determined by the characteristics of that surface on a physical scale on the order of the wavelength of the acoustic signal. Although a specular reflection is the result of a wave reflecting off a flat, non-absorptive material, a non-specular reflection can be caused by a number of surface characteristics such as irregularities in the shape or absorption coefficient (and therefore acoustic impedance). In order to evaluate the specific qualities of assorted diffusion properties, various surface characteristics are discussed.

1.2.2.1 Irregular surfaces

The natural world is comprised of very few specular reflectors for light waves – even fewer for acoustic signals. Until the development of artificial structures, reflecting surfaces were, in almost all cases, irregularly-shaped (with the possible exception of the surface of a very calm body of water). As a result, natural acoustic reflections are almost always diffused to some extent. Early structures were built using simple construction techniques and resulted in flat surfaces and therefore specular reflections.

For approximately 3000 years, and up until the turn of the 20th century, architectural trends tended to favour florid styles, including widespread use of various structural and decorative elements such as fluted pillars, entablatures, mouldings, and carvings. These random and periodic surface irregularities resulted in more diffused reflections according to the size, shape and absorptive characteristics of the various surfaces. The rise of the “International Style” in the early 1900’s (Nuttgens 1997) saw the disappearance of these largely irregular surfaces and the increasing use of expansive, flat surfaces of concrete, glass and other acoustically reflective materials. This stylistic move was later reinforced by the economic advantages of these design and construction techniques.

1.2.2.2 Maximum length sequence diffusers

The link between diffused reflections and better-sounding acoustics has resulted in much research in the past 30 years on how to construct diffusive surfaces with predictable results. This continues to be an extremely popular topic at current conferences in audio

and acoustics with a great deal of the work continuing on the breakthroughs of Schroeder. In his 1975 paper, Schroeder outlined a method of designing surface irregularities based on maximum length sequences (MLS) (Golomb 1967) which result in the diffusion of a specific frequency band. This method relies on the creation of a surface comprised of a series of reflection coefficients alternating between +1 and -1 in a predetermined periodic pattern.

Consider a sound wave entering the mouth of a well cut into the wall from the concert hall as shown in Figure 1.9.

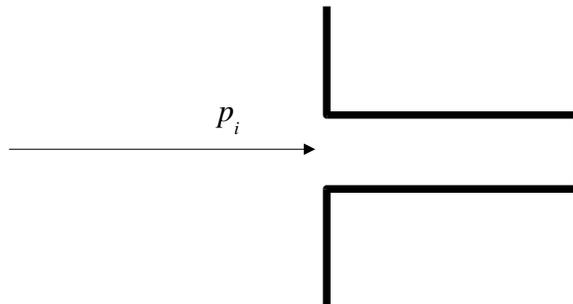


Figure 1.9 Pressure wave incident upon the mouth of a well.

Assuming that the bottom of the well has a reflection coefficient of 1, the reflection returns to the entrance of the well having propagated a distance equalling twice its depth d_w , and therefore undergoing a shift in phase relative to the sound entering the well. The magnitude of this shift is dependent on the relationship between the wavelength and the depth according to Equation 1.18.

$$\phi = 4\pi \frac{d_n}{\lambda} \tag{1.18}$$

where ϕ is the phase shift in radians, d_n is the depth of the well and λ is the wavelength of the incident sound wave.

Therefore, if $\lambda = 4 d_n$, then the reflection will exit the well having undergone a phase shift of π radians. According to Schroeder, this implies that the well can be considered to have a reflective coefficient of -1 for that particular frequency, however this assumption will be expanded to include other frequencies in the following section.

Using an MLS, the particular required sequence of positive and negative reflection coefficients can be calculated, resulting in a sequence such as the following, for N=15 (Schroeder 1975):

- + + - + - + + + + - - - + -

This is then implemented as a series of individually separated wells cut into the reflecting surface as is shown in Figure 1.10. Although the depth of the wells is dependent on a single so-called *design wavelength* denoted λ_0 , in practice it has been found that the bandwidth of diffused frequencies ranges from one-half octave below to one half octave above this frequency (Schroeder 1975). For frequencies far below this bandwidth, the signal is typically assumed to be unaffected. For example, consider a case where the depth of the wells is equal to one half the wavelength of the incident sound wave. In this

case, the wells now exhibit a reflective coefficient of +1; exactly the opposite of their intended effect, rendering the surface a flat and therefore specular reflector.

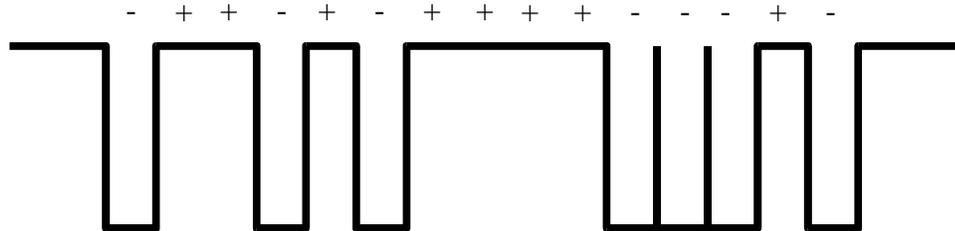


Figure 1.10 MLS Diffuser showing the relationship between the individual wells cut into the wall and the MLS signal.

The advantage of using a diffusive surface geometry based on maximum length sequences lies in the fact that the power spectrum of the sequence is flat except for a minor dip at DC (Schroeder 1975). This permits the acoustical designer to specify a surface that maintains the sound energy in the room through reflection while maintaining a low interaural cross correlation (IACC) through predictable diffusion characteristics which do not impose a resonant characteristic on the reflection. The principal disadvantage of the MLS-based diffusion scheme is that it is specific to a relatively narrow frequency band, thus making it impractical for wide-band diffusion.

1.2.2.3 Schroeder diffusers

The goal became to find a surface geometry which would permit designers the predictability of diffusion from MLS diffusers with a wider bandwidth. The new system was again introduced by Schroeder in 1979 in a paper describing the implementation of

the *quadratic residue diffuser* or *Schroeder diffuser* (Schroeder 1979) – a device which has since been widely accepted as one of the de facto standards for easily creating diffusive surfaces. Rather than relying on alternating reflecting coefficient patterns, this method considers the wall to be a flat surface with varying impedance according to location. This is accomplished using wells of various specific depths arranged in a periodic sequence based on residues of a quadratic function as in Equation 1.19 (Schroeder 1979).

$$s_n = n^2, \text{mod}(N) \tag{1.19}$$

where s_n is the sequence of relative depths of the wells, n is a number in the sequence of non-negative consecutive integers $\{0, 1, 2, 3 \dots\}$ denoting the well number, and N is a non-negative odd prime number.

For example, for modulo 17, the series is

0, 1, 4, 9, 16, 8, 2, 15, 13, 13, 15, 2, 8, 16, 9, 4, 1, 0, 1, 4, 9, 16, 8, 2, 15 ...

As may be evident from the representation of this series in Figure 1.11, the pattern is repeating and symmetrical around $n = 0$ and $N/2$.

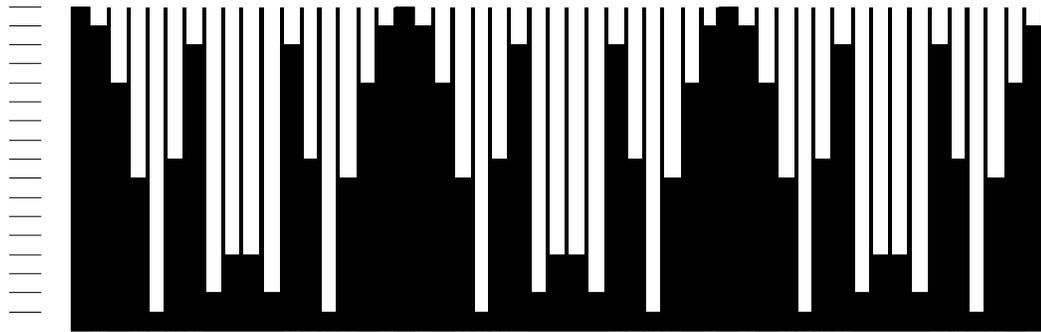


Figure 1.11 Schroeder diffuser for $N = 17$ following the relative depths listed above.

The actual depths of the wells are dependent on the design wavelength of the diffuser. In order to calculate these depths, Schroeder suggests the function

$$d_n = s_n \frac{\lambda_o}{2N} \quad (1.20)$$

where d_n is the depth of well n and λ_o is the design wavelength (Schroeder 1979).

The widths of these wells w should be constant and small compared to the design wavelength (no greater than $\lambda_o/2$; Schroeder suggests $0.137 \lambda_o$). Note that the result of Equation 1.20 is to make the median well depth equal to one-quarter of the design wavelength. Since this arrangement has wells of varying depths, the resulting bandwidth of diffused sound is increased substantially over the MLS diffuser, ranging approximately from one-half octave below the design frequency up to a limit imposed by $\lambda > \lambda_o/N$ and, more significantly, $\lambda > 2w$ (Schroeder 1979).

The result of this sequence of wells is an apparently flat reflecting surface with a varying and periodic impedance corresponding to the impedance at the mouth of each well. This surface has the interesting property that, for the frequency band mentioned above, the reflections will be scattered to propagate along predictable angles with very small differences in relative amplitude.

Each of the wells in a quadratic residue diffuser can be considered as a simple quarter-wavelength resonator consisting of a pipe which is open on one end and terminated by a known impedance at the other. The acoustical impedance Z_0 of an acoustical plane wave propagating inside a circular pipe with a diameter less than the wavelength of the sound wave in a direction along its length can be calculated using Equation 1.21 (Kinsler 1982). This impedance differs from the characteristic impedance of air since the space is now bounded by the well diameter.

$$Z_0 = \frac{\rho_o c}{S_n} = \frac{z_o}{S_n} \quad (1.21)$$

where S_n is the cross sectional area of the pipe and therefore $S_n = \pi r_n^2$ where r_n is the radius of the pipe.

The impedance at the entrance of a pipe (at $x = 0$) closed on the opposite end is shown in Equation 1.22 (Kinsler 1982).

$$Z_n = \frac{\rho_o c}{S_n} \frac{z_d + j \frac{\rho_o c}{S_n} \tan(k d_n)}{\frac{\rho_o c}{S_n} + j z_d \tan(k d_n)} \quad (1.22)$$

where k is the wave number and is expressed in Equation 1.23., z_d is the impedance of the cap closing the bottom of the pipe, and d_n is the depth of the pipe.

$$k = \frac{\omega}{c} = \frac{2\pi f}{c} = \frac{2\pi}{\lambda} \quad (1.23)$$

If the closed end is given an impedance of $z_d = \infty$ then Equation 1.22 simplifies to Equation 1.24 (Kinsler 1982).

$$Z_n = -j \frac{\rho_o c}{S_n} \cot(k d_n) \quad (1.24)$$

Note that the acoustic reactance component of the impedance becomes 0 when $\cot(k d_n)$ is set to equal 0 as is evident in Equation 1.25 (Kinsler 1982).

$$k d_n = (2h - 1) \frac{\pi}{2} \quad (1.25)$$

where $h = \{1, 2, 3 \dots\}$

Manipulating Equation 1.25, we arrive at Equation 1.26 (Kinsler 1982).

$$f = \frac{(2h - 1) c}{4 d_n} \quad (1.26)$$

This produces the frequency pattern of a quarter-wavelength resonator where the

fundamental resonance is $\frac{c}{4d_n}$ when $h = 1$.

As can be seen in Figure 1.12, when this calculated frequency-dependent impedance is plotted for a pipe with a length of 8.6 cm, a diameter of 4.7 cm (and therefore a design frequency of 1000 Hz), and a cap with an infinite acoustic impedance, a number of the reflective characteristics of the well are evident. Firstly, it should be noted that there is no real component in the acoustic impedance. This is evident in Equation 1.24 which demonstrates that, when the closed end has an infinite acoustic impedance, the input impedance of the open end of the pipe contains only an imaginary component. As a result, the graph plots only the acoustic reactance. Secondly, the impedance response is periodic in the frequency domain with a period of two times the design frequency. This is a result of the reflected signal undergoing an increasing phase shift with frequency from 0 to π radians for an increase from 0 Hz to the design frequency. As the frequency increases to the second harmonic of the design frequency, the phase shift continues to increase, albeit effectively from $-\pi$ radians to 0, resulting in an infinite impedance at twice the design frequency.

1.2.2.3 Schroeder diffusers

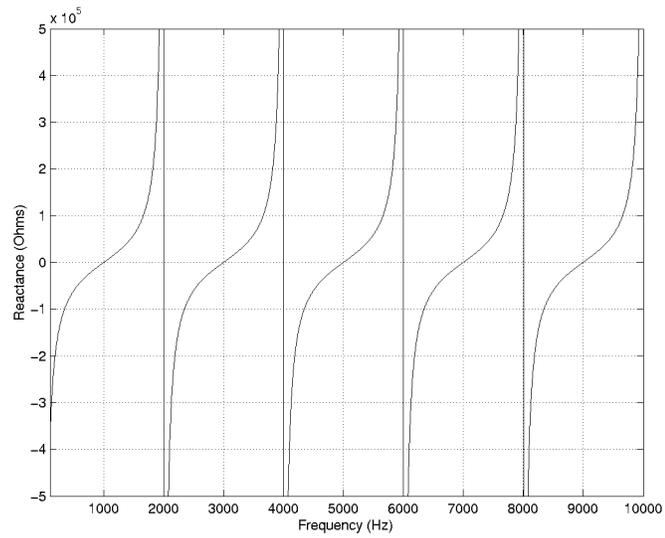


Figure 1.12 Acoustic reactance vs. frequency for the input of a diffuser well with a design wavelength of 1 kHz (depth = 8.6 cm, width = 4.17 cm) and a bottom with an infinite acoustic impedance. Note that, in this particular case, the acoustic resistance at all frequencies is 0 Ω .

Since the wells in a quadratic residue diffuser are different lengths, the resonant peaks in this local impedance response and therefore the “frequency-dependent reflective coefficient response” vary accordingly, however the characteristic shape of the local impedance curve remains.

1.3 Perceptual significance of acoustic characteristics

The perception of our surrounding environment is the product of a number of sensory input channels which tend to be dominated by visual and auditory stimuli. Although there is no question that the awareness of our environment and the objects contained within it is predominantly based on optical recognition, it can be argued that we in fact, receive more information on more parallel “streams” (Bregman 1990) through our hearing than our sight. For example, while our field of vision is limited to a relatively narrow angle, and then only while our eyes are open, we are continually receiving aural information from sound sources and reflections from all directions simultaneously. This larger capacity to process incoming information applies not only to spatial characteristics, but to temporal ones as well. While we can be misled into believing that a series of still pictures presented at a rate of 24 frames per second is the equivalent of smoothly continuous movement, we require a minimum of approximately 40,000 samples per second to mimic realistic-sounding auditory events. Despite these facts, we accept poor acoustical conditions in a concert hall far more readily than we would poor lighting conditions in a fine art museum. A more typical example of our indifference is the mass acceptance of poor audio quality during an entire movie, but rapid and frequently vocal intolerance for poor focus of the image on the screen.

One of the more interesting characteristics of the auditory system is demonstrated by what is commonly known as the “cocktail party effect” (Cherry 1953). This is the ability of the brain to focus on a single component – typically perceived as the direct sound from a

source – within a large group of parallel contributing and conflicting signals. Although this allows us to concentrate on a single voice in a crowded room or a musical instrument in a reverberant space, it does not preclude the perceptual system's simultaneous reliance on the other components to provide us with information regarding both the sound source and its environment. In fact, in many cases, it can be easily proven that the intensity of the direct sound received from a source is far lower than the sum of the intensities received from all other sources of auditory information in a given situation (Zwicker 1999). This sum of extraneous sounds is comprised of various related and unrelated streams which include the reflections and reverberation generated by the sound source in question. These can include supporting sounds such as adjacent violins in a string section or conflicting sounds such as crinkling candy wrappers belonging to a fellow audience member.

1.3.1 Quantitative percepts

There are a number of quantitative attributes of a sound source and its surroundings which can be determined using only the auditory system. Some of these attributes, such as the source's direction or angular location do not rely heavily on additional information from surrounding reflective surfaces. Others, such as the distance to the source as well as the location and characteristics of reflecting surfaces are heavily dependent upon the relationship between the direct and reflected sound waves at the listening position.

In a perfectly anechoic environment there are no reflections, reverberation or resonances other than those produced by the sound source and the listeners themselves. In such a situation, we are presented with very little quantitative information that would normally

1.3.1 Quantitative percepts

provide us with indications of such cues as distance to the sound source and the dimensions of the surrounding enclosure. Interaural amplitude and time delay (or phase) differences provide us with information regarding angular location, allowing us to determine the source's position with varying degrees of accuracy ranging from $\pm 1^\circ$ to $\pm 20^\circ$ depending on the direction and plane of movement (Blauert 1997).

In specific cases, it may be possible to determine the distance to a sound source in this anechoic environment. For familiar sources, the simple level of the received signal can provide an indication of distance (Moore 1997). Also, when the sound source and the listener are separated by large distances, there are high-frequency attenuation cues caused by the absorptive effects of the air, (Coleman 1962)(Coleman 1963) however, without knowing the frequency response and level of the sound source at its location, as is the case for unfamiliar sound sources, we are unable to use these components for determining distance.

As the number and complexity of surrounding reflective surfaces increases, the listener is presented with more information to process regarding both the sound source and the environment which contains it. It has been demonstrated that we derive information regarding the location of sound sources from both the pattern of early reflections and the ratio between the direct and reverberant sound powers. In the former case, the information is dependent upon the relative levels, times of arrival and locations of the direct sound and its individual reflected copies at the listening position. If these reflections occur within approximately 30 ms of the direct sound, we perceive them as

being components of the direct sound (Wallach 1949), however their importance for distance perception has been proven. Von Békésy (1960) showed that the apparent distance to the sound source could be modulated by altering only the reflection patterns. Mershon and Bowers (1979) proved that these cues provide information regarding both the relative and the absolute distance to the sound source and are used by listeners for familiar and unfamiliar sounds in familiar and unfamiliar environments. It has been shown that human listeners are able to use the direct-to-reverberant energy ratio to estimate the distance to a sound source. Flanagan and Taylor (1999) noted listeners' tendencies to underestimate the distance to sources with limited radiation angles and therefore increased direct-to-reverberant energy ratios. The same listeners were able to estimate much more accurately the location of diffusely radiating omnidirectional sources with more predictable direct-to-reverberant energy ratios.

The characteristics of the early reflections also have a measurable effect on the sound field at the listener's position. These range from simple frequency response differences to an altered power distribution over the listening area.

1.3.1 Quantitative percepts

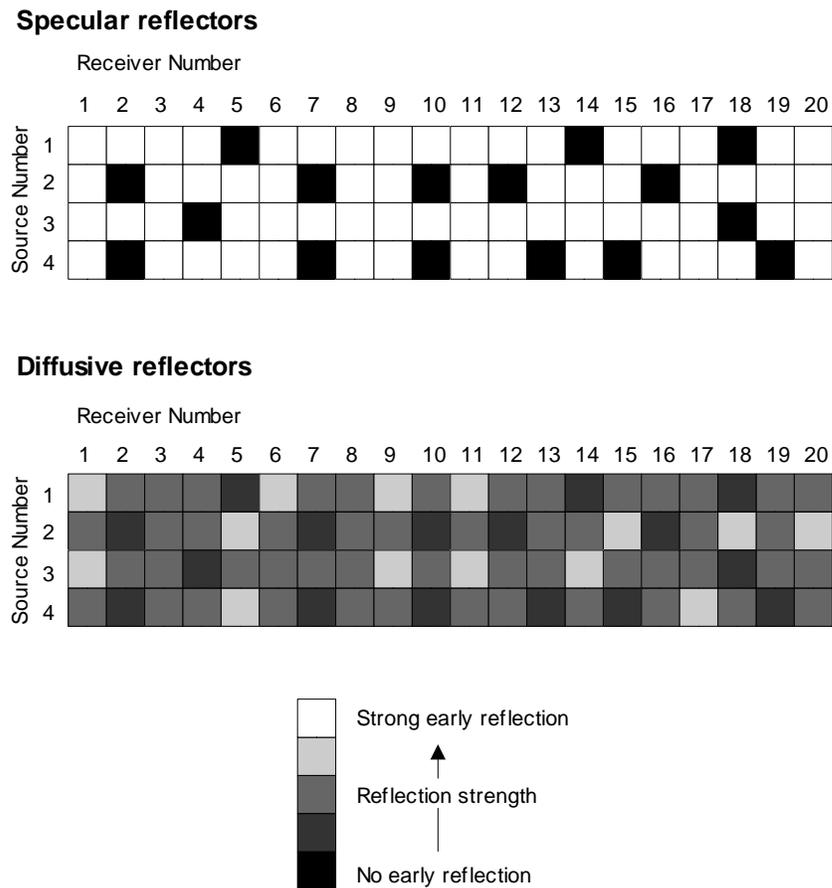


Figure 1.13 The relationship between early reflection power received at the listening position and reflective surface characteristics for 4 sound sources and 20 receivers in a hypothetical space. Adapted from Dalenbäck (1994)

The specific effect of diffused vs. specular reflections on the power received at different locations in an enclosure has been discussed by Dalenbäck (1994). Figure 1.13 provides an excellent and intuitive way of considering the distribution of power in reflections to various locations in the audience of a performance space. Each graph shows an intensity plot of strength of reflected energy for various combinations of instrument and audience member locations. The top plot illustrates highly specular reflective surfaces and

1.3.1 Quantitative percepts

demonstrates that, while most audience members receive a high level of reflected energy, some source / receiver combinations, due to relative locations, result in no reflected energy at the receiver's position. In the case of diffusive reflectors, plotted in the bottom graph, note that the highest calculated level of energy is lower, however all combinations of source and receiver result in at least some reflected energy at the listeners' locations. While this situation is not immediately evident in an acoustical context, it is easily conceivable in a room of irregular geometry with mirrored walls. Consider that there would be a number of light source and viewer locations in such a room in which all reflections seen by the viewer are the result of higher-order reflections.

1.3.2 Qualitative percepts

In addition to their importance for listeners' evaluations of quantitative aspects of their environment such as sound source and boundary location, the particular characteristics of individual reflections have a heavy influence on the perceived aesthetic quality of the sound sources and of the acoustic environment. In a 1974 comparison of a number of European concert halls, it was determined that a lower interaural coherence caused by more diffused reflections correlated with a greater preference of listeners (Schroeder 1974). More recently, Hann and Fricke demonstrated that there is a high degree of positive correlation between the *Surface Diffusivity Index* (SDI), a measure of reflecting surface roughness based on relatively simple visual inspection, and the *Acoustic Quality Index* (AQI) in a large number of the world's recognized concert halls (Hann 1993). In their words:

Surface diffusivity appears to be largely responsible for the difference between halls which are rated as excellent as opposed to those rated as good or mediocre.

Beranek (1996) argues that this statement is “overly inclusive,” but does not dispute that the diffusive qualities of reflective surfaces are among the more important characteristics which determine the acoustical quality of a concert hall, stating that

Diffusivity is an architectural feature that must not be underestimated.

This statement is not a modern concept by any means. It has been known for at least 100 years that irregularities in the reflective surfaces have a positive effect on sound quality (Beranek 1996). This aural aesthetic peacefully coexisted with the influence of Greek and Roman architectural styles in the design of older structures, resulting in the widespread use of decorative topologies such as relief ornamentation, sculptures and columns which together act as excellent structures for the dispersion of reflected sound waves (D’Antonio 1995). Regrettably, the rise of the “International Style” at the turn of the last century and modern economic restrictions result in reduced time frames for construction, increasing building costs and higher seating capacities. The result of this has been that the design of modern spaces is typically characterized by large, flat surfaces composed of concrete, plaster and drywall with little attention paid to the necessity of boundaries with diffusive properties (D’Antonio 1995). While this design aesthetic satisfies both minimalist architects and accountants, it imposes great detriments on acousticians, musicians and audiences.

There are a number of physically measurable effects of increased diffusion in a reverberant space which can be correlated with preferences of listeners. Schroeder noted the decreased IACC which results from greater surface diffusivity (Schroeder 1974). A number of researchers since then have found correlations between an increased sense of spaciousness (and therefore higher degrees of preference) and lower IACC's (Schroeder 1974) (Hidaka 1995). This decreased IACC for transient program material is the product of a stochastic reflecting surface producing a more complex impulse response. For steady state low frequencies, there is a decreased prominence of characteristic room resonances (Angus 1999), thereby reducing interaural phase similarities. In addition, diffusion decorrelates the various reflections both with the direct sound and with each other, thus reducing undesirable resonances at the listening position caused by comb-filtering effects. As a result, it is possible using diffusive reflectors to maintain acoustic energy in the enclosure over a longer time period in its impulse response without causing the audible interference generated by specular reflections.

To summarize, in most cases, it is necessary to provide listeners with early reflections in order to provide cues both for location of the sound source and the size of the enclosure, however, these reflections must be diffused in order to maintain a high aesthetic quality and to ensure an even distribution of energy to all listeners.

1.4 Significance of acoustics simulation

Up until the past century, music was composed with the intention that it be heard in spaces with numerous diffused reflections and longer reverberation times. Current pop and electroacoustic music relies very heavily on the application of synthetic reverberation to recordings with little to no natural acoustic space. In contrast, today's typical real environments, both residential and public, are generally characterized by a small number of predominantly specular reflections from flat walls and short reverberation times due to highly absorptive furnishings and surface coverings such as carpet and ceiling tiles. As a result, it is common practice to add extra synthetic reflections and reverberation to our recordings. Irrespective of usage, artists, producers and recording engineers simultaneously demand both very high aesthetic quality from and detailed control over these synthesized environments, tailoring the various attributes of the virtual space according to the specific requirements of the project.

Assuming that we are able to develop a synthetic acoustical model that rivals the perceptual quality of real spaces, there are a number of advantages for using these systems.

1. **Flexibility and repeatability:** Whereas it is time consuming and usually expensive to modify the acoustical characteristics of a real space, either with incorporated systems such as movable panels and curtains or temporary arrangements, modification of the characteristics of virtual acoustic spaces is simply a matter of changing parameter values on a display.

2. **Availability:** Most classical recording engineers and producers are aware of the value of a recording space that not only possesses desirable acoustical characteristics, but is available for use when required. Of course, a synthesized acoustic environment does not suffer from availability difficulties, barring events such as power failures.
3. **Potential for improvement:** Unless the reverberation engine is based entirely on a physical model, it is possible to achieve acoustical characteristics in a synthetic environment which would be physically impossible in a real one. This is a technique commonly used in pop and electroacoustic music where surreal environments are integral to the musical style. In the case of classical styles, it is conceivable that acoustical environments could be created to match the specific musical style or performance, thus fine-tuning the acoustics with a result better than that which is practically available in real spaces.
4. **Predictive software:** The past 15 years have seen a growing use of virtual reality systems in architectural planning and design. When clients are able to “walk through” their buildings before construction has begun using three-dimensional representations generated from the existing CAD files, modifications to the design are quite inexpensive to undertake, particularly when compared with the costs of demolition after the structure is completed. This advantage is true, not only of spatial characteristics of the enclosure, but of its acoustical characteristics. The ability to hear a musical performance played “in” a concert hall that has not yet

been built would allow designers and clients to modify and improve the design, thus avoiding disasters such as that which occurred in the construction of New York City's Philharmonic Hall in 1962 (Glanz 2000). (It should be noted that, generally speaking, acoustics simulation intended for production of music recording would be significantly different from that used for predicting real acoustical environments. While there would be some overlap of various parameters, the requirements of recording production frequently demand physically impossible spaces as was discussed above.)

1.5 Historical review of synthetic digital reverberation algorithms

Since the application of digital signal processing to audio signals, there have been many attempts to create an impression of space using various algorithms. For the past 40 years, numerous researchers and commercial developers have devised multiple strategies for synthesizing virtual acoustic environments with varying degrees of success. Different methods have proven to be more suited to particular applications which can be divided into two major fields:

1. commercial classical music production and film and television post-production
2. commercial pop music production, electroacoustic music composition (including computer music) and computer games.

In the former case, synthetic reverberation is typically used to mimic real acoustic environments, generally with the intention of deceiving listeners into believing it is an actual space. In the latter case, the synthetic environments are generally intended to appear to be exactly that – synthetic – acoustics which are physically impossible in the real world, essentially resulting in the processing being used more as an instrument than an environment. In either instance, the reverberation must be aesthetically acceptable in order to enhance and complement rather than detract from the performance it accompanies.

Unfortunately, as is pointed out by Gardner (1992a), since a large part of the development of synthetic reverberation algorithms has been carried out by corporate concerns for

commercial products, many secrets regarding their algorithms are not shared. Companies that rely on the sale of such products maintain superior product quality by ensuring that their procedures are not used by competitors, thus knowledge is hoarded rather than distributed, consequently slowing progress in the field. As a result, many specifics regarding current commercially-developed algorithms can only be either surmised or reverse-engineered.

Various algorithms and approaches to acoustics simulation have been developed since the advent of digital signal processing in the audio and film industries. These can be loosely organized into three groups based on their basic procedural characteristics: 1) *tapped recirculating delay models*, 2) *geometric models* and 3) *stochastic models*.

1.5.1 Tapped recirculating delay model

Most modern digital reverberation engines use a collection of multitap delays, recursive comb filters and allpass filters connected in series and parallel with various combinations of parameter configurations to generate artificial acoustic spaces. Since these systems are based on assemblies of simple tapped delays with feedback and gain control they fall under the general term of “tapped recirculating delay” or TRD models (Roads 1996). This has been the standard method of generating artificial reverberation for approximately 40 years.

The first widely accepted method of creating synthetic reverberation using digital signal processing is attributed to M.R. Schroeder who, in 1961 and 1962, published two seminal

1.5.1 Tapped recirculating delay model

papers describing procedures using simple series and parallel arrangements of recursive comb and allpass filters labelled “unit reverberators” (Schroeder 1961) (Schroeder 1962) (see Figure 1.14) Despite the fact that these models used hours of computation time on mainframe computers, they remained the algorithms of choice and were used extensively at centres such as Stanford University and IRCAM for over a decade (Moorer 1979).

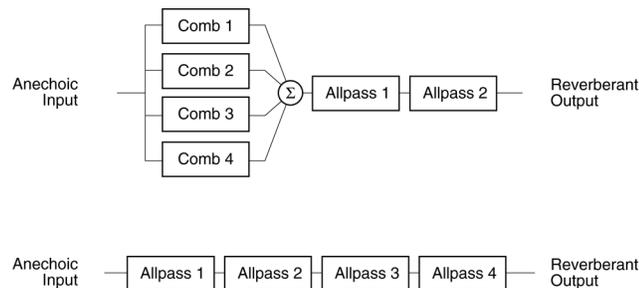


Figure 1.14 Schroeder’s original two reverberation algorithms based on unit reverberators

In 1970 Schroeder suggested an improvement to his system with the inclusion of early power in the form of discrete reflections which will be discussed in the subsequent section (Schroeder 1970). While the Schroeder algorithms were universally accepted they do exhibit a number of problems outlined in 1979 by James Moorer. These complaints centre around the lack of perceptual transparency of the system, primarily a result of the excessive simplicity of the unit reverberators. Moorer specifies three major problems with the system (Moorer 1979).

1. A “lag” in the reverberant decay buildup, dependent on the order of the system.
Unlike the case in real acoustics which appear to start with a dense reverberant

1.5.1 Tapped recirculating delay model

field and decay exponentially, the Schroeder algorithm tends to have an audibly slower attack envelope.

2. The “smoothness” of the decay is highly dependent on the relationship of the various gain and delay parameter settings of the unit reverberators. Despite the use of prime relationships between various delay times, “ragged-sounding” decays were easy to achieve.
3. An audible “ringing” effect is frequently the result of the algorithm, typically related to the resonant frequencies caused by the delay time relationships in the unit reverberators.

Moorer (1979) suggested a number of improvements on Schroeder’s original algorithm including the introduction of two new unit reverberators, termed an “oscillatory all-pass” and an “oscillatory comb,” and the use of low-pass filtering in the feedback loops both to smear the time response of discrete echoes and to simulate the absorption of high frequency signals in air for the reverberant field.

1.5.2 Geometric model

The TRD model relies heavily on the perceptual attributes of the unit reverberators in the attempt to mimic real acoustics. While the use of various combinations of recursive comb and allpass filters approximates the diffuse decay of a reverberant space, there is no effort to replicate precise characteristics of physical spaces. Although this system can be used to provide an adequate simulation of the later components in a reverberant decay, it can only

be expected to provide a rough approximation for critical applications in spite of Schroeder's claim in 1962 that the output from such a system, in some regards, was perceptually "indistinguishable" from recordings made in real spaces.

In 1970, the TRD system was improved dramatically with Schroeder's inclusion of a multitap delay which provided simulated early reflections (Schroeder 1970). Over the past 30 years, this system has undergone numerous improvements, (Gardner 1992b) (Christensen 1999) thanks largely to increases in complexity with increased processing power. Despite the fact that this model was the first system to be used to synthesize artificial acoustic environments, it remains the most commonly-used model in even the newest commercial devices. Its initial development relied on the efficiency of the TRD algorithms and the use of recursive elements to model reverberation perceptually. While this system does not provide an accurate representation of the real characteristics of a reverberant space, its efficiency and perceptual adequacy lie at the heart of its continued acceptance.

There are two principal methods used in calculating the real temporal and spatial characteristics of the reflection pattern of an enclosure, each providing advantages and disadvantages. These are the *ray-trace model* and the *image model*. Two additional methodologies are derived from these and are known as the *cone trace model* (also known as *beam tracing*) and the *hybrid method*.

1.5.2.1 Ray-trace model

Schroeder's intention in the addition of a multitap delay to the TRD system was to simulate the early reflection pattern of a real enclosure. In order to determine the appropriate delay times, gains and panning for these delays, he chose to use a two-dimensional model of a real room, calculating the paths of 300 sound rays propagating outwards from the sound source towards the walls (Schroeder 1970). Assuming linear response characteristics both from the air and the reflective surface, he subsequently calculated a series of delays to simulate the early reflection patterns for a given source and listener location and specified room geometry. This delay cluster was then used to provide the early reflection pattern of the room, with the TRD model being used to provide the desired reverberant characteristics.

This is a simple example of the system which has come to be known as *ray-tracing*. In it, the expanding wavefront is considered as a number of rays of energy, typically equally spaced angularly around the sound source. As each ray propagates outwards, it will meet a reflecting surface and subsequently be re-directed into the enclosure at a new angle. This procedure continues until either a pre-determined limit on the order number of the reflection is reached or the ray encounters the receiver. The appropriate time and gain of each ray at the receiver is therefore calculated and stored, with the collection of all incoming rays resulting in the calculated impulse response of the room. In more sophisticated systems, the absorptive characteristics of each reflecting surface can be

1.5.2.1 Ray-trace model

included by applying an appropriate filter to the impulse for each corresponding reflection.

This system is an excellent method for calculating the impulse response of a room with numerous surfaces with irregular geometry since the local interaction of each ray with the surrounding environment is considered throughout its entire propagation. Consequently, angular calculations are made in series and are therefore relatively simple. As a result it is the system most widely used in software developed for acoustics prediction (Dalenbäck 1994). One principal danger with this system is that, unless a large number of rays are used, it is possible that low-order reflections will be omitted since they will occur between angles. This is because the radiation patterns of sound sources are calculated on quantized angles instead of continuous space. Consequently, it is possible that the location of the receiver does not correspond to any of the emitted rays for some reflections.

1.5.2.2 Image model

While the ray trace model provides an excellent and efficient method of calculating the spatial and temporal locations of reflections in an enclosure with irregular geometry, it requires a very large number of calculated rays in order to assure that all early reflections are found. In cases where only the low-order reflections are required, the image model provides a more efficient method of calculation for specular reflective surfaces. In this model, the sound source is effectively duplicated to multiple locations outside the enclosure. These locations correspond to the virtual locations of sound sources which, in

a free field, would result in the same delayed pressure wave as a perfect specular reflection. Figure 1.15 illustrates this concept. The figure on the left demonstrates the direct sound and first reflection using a simple ray trace model. In it, the reflection is calculated to occur at a particular location on the reflecting surface. In the figure on the right, the image model is used where the sound source is duplicated on the opposite side of the reflecting surface. The incoming pressure wave from this duplicated source exactly matches the pressure wave of the reflection in the diagram on the left.

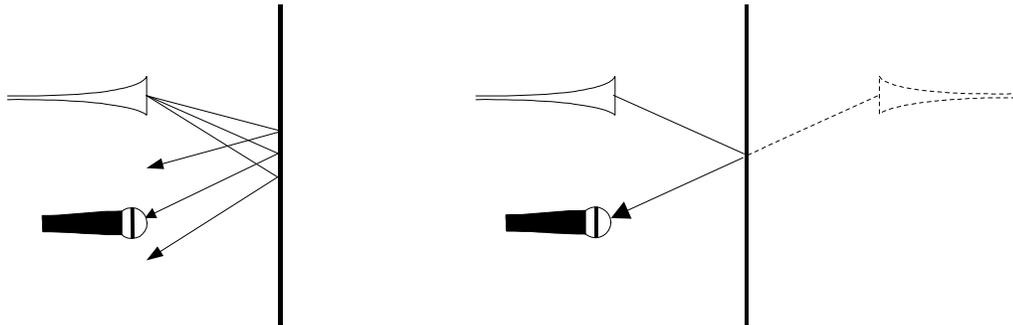


Figure 1.15 Simple ray trace model (left) vs. image model (right)

This model was first proposed for the building of impulse responses of digitally synthesized acoustic environments in 1979 by Allen and Berkley (1979) however, the method was used by many other researchers previous to this to model specular reflective surfaces of a rectangular enclosure (Eyring 1930) (Gibbs and Jones 1972) (Berman 1975). In this paper, they suggested finding the total propagation delays of a reflection using this model, and subsequently rounding the delay to the nearest sample in the synthetic impulse response. This system could be repeated for any number of reflections with the resulting

impulse response used as an FIR filter through which anechoic recordings would be convolved.

Peterson (1986) suggested an improvement to this system in the form of interpolated delays. Whereas Allen and Berkley assumed that a maximum temporal quantization error, caused by rounding off delay values, of $\frac{\pm T_s}{2}$ was tolerable, (where T_s is the sampling period) Peterson suggested that this corresponded to an unacceptable interchannel phase error as will be discussed in reference to the proposed diffusion simulation system in Chapter 4. This error is eliminated by using interpolated rather than quantized delay values.

The advantages of the image model system lie in the efficiency of its method of calculation for specific reflections. The impulse responses of individual reflections are calculated as desired, and thus low-order reflections are provided by the system very quickly. The principal disadvantage of the system is its difficulty in processing irregular geometries, particularly for higher order reflections. Whereas the ray trace model computes the various reflections of a single ray in series, the image model does so in parallel – all reflections of a given ray must be known simultaneously in order to return a result (Borish 1984). Consequently, this model is used in systems where the enclosure is constructed of relatively simple shapes and only lower-order reflections are required.

1.5.2.3 Cone trace model

The cone trace method was developed in an attempt to maintain the simplicity of the ray trace model for higher-order reflections, while ensuring that lower-order reflections were not omitted due to quantization of the angular directions of the sound radiation. In this procedure, rather than using discrete rays emanating from the sound source at specific angles, a series of H overlapping cones are calculated with a gain function which is dependent on the relationship between the angular spread of the cone γ_0 (derived using Equation 1.27) and the location of the receiver at angle γ . This gain, G_γ is calculated using Equation 1.28 (Maercke 1986).

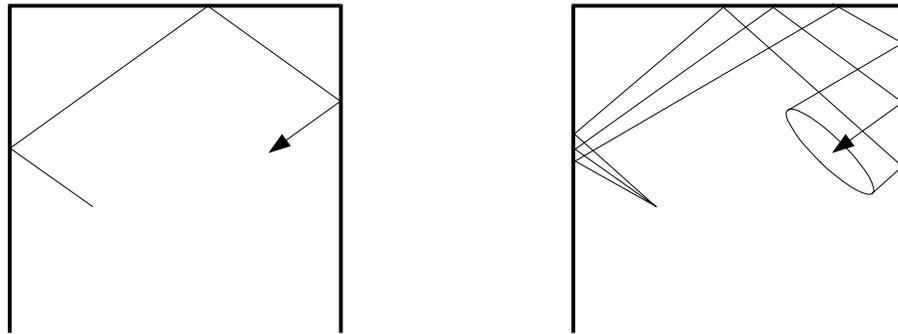


Figure 1.16 Simple ray trace method (left) vs. cone trace method (right)

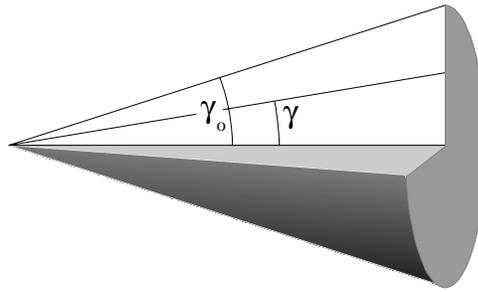


Figure 1.17 Angles used for calculating gain within cone (see Equation 1.27). In order to simplify the diagram, the reflected cone shown in Figure 1.16 has been shown without the reflections.

$$\gamma_0 = 1.05 \sqrt{\frac{4\pi}{H}} \quad (1.27)$$

$$G_y = \cos^2\left(\frac{\gamma}{\gamma_0}\right) \quad (1.28)$$

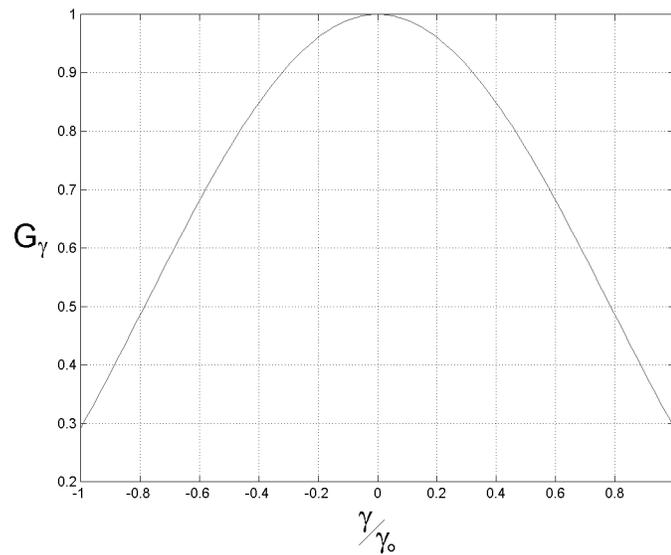


Figure 1.18 G_γ vs. angle ratios for cone trace model

1.5.2.4 Hybrid method

Another approach which combines the simplicity of the ray trace model with the accuracy of the image model is the hybrid method. This system is simply a combination of the two, using the image model for lower-order reflections and ray tracing for the higher orders (Vörlander 1989).

1.5.3 Physical models

While the ray-trace and image models use mathematical descriptions of real acoustic spaces, they do not necessarily mimic the behaviour of an acoustic wavefront in those spaces. This is due largely to the fact that each considers only the propagation speed and direction of the expanding wavefront, and does not take into account such factors as the diffusion or diffraction (Kleiner 1993). As a result, although these models will provide

excellent models of early specular reflection patterns, they do not produce accurate impulse responses for high orders of reflections. This is because the spatial spread of quantized radiation angles increases with distance from the sound source, and thus the order of the reflection. Consequently, it is likely that many reflections will not be recorded at the receiver's position. In order to achieve a higher level of accuracy, a model which is based on a larger set of physical rules is required. In this so-called *physical model*, the system is equipped with equations that describe the mechanical and acoustic behaviour of the various components of the physical system being modeled. In theory, the result is a system that mimics the behaviour of the physical counterpart (Roads 1996).

The term “physical model” is widely-used in many fields to describe various systems. In music technology-based applications, a physical model is one where mathematical models of physical acoustics are used to produce the characteristics of a resonant instrument or enclosure (Roads 1996). This mathematical model can take various forms, however, the typical implementation involves a recursive delay including a filter in the feedback path, an example of which is the ubiquitous Karplus-Strong plucked string algorithm (Karplus 1983).

Two general methods of physical models used in auralization and predictive acoustics software are known as the *boundary element method*, and the *finite element method*.

1.5.3.1 Boundary element method

Using the boundary element method, surfaces are subdivided into discrete components, each with a particular set of reflective characteristics (Kleiner 1993). Each of these components is considered to be a sound source, re-radiating power into the enclosure with an individual contribution to the whole room impulse response. This methodology will be explained further in Chapter 3, since it forms the foundation of the method of wall construction in the model proposed in this dissertation.

One of the principal problems with the boundary element method is the large number of calculations required in order to build a complete impulse response. This is particularly true with higher order reflections, since the increase in the number of interacting elements is exponentially proportional to the order of reflection.

1.5.3.2 Finite element model

The boundary element method is limited in that it only considers the characteristics of the surfaces which define a space while neglecting the behaviour of elements within the space itself. The finite element model corrects for this omission, subdividing the entire room into a collection of interconnected discrete elements arranged in a mesh. In this manner, the entire space is modelled, albeit with a rather high computational cost. One notable example of the finite element method is the *digital waveguide mesh*.

1.5.3.3 Digital waveguide mesh

Various systems have been developed and proposed which use the concept of digital waveguides to simulate the resonant and reverberant characteristics both of room acoustics and, on a smaller scale, of instrument acoustics. Initially proposed for room models by Crawford in 1968 (Roads 1996), digital waveguides use bidirectional delay lines (and therefore recursion) to simulate the characteristics of acoustic waveguides with considerable efficiency. A notable development in the field of room acoustics modelling was the extension of the digital waveguide into a multi-dimensional mesh. This mesh is comprised of a number of discrete digital waveguides that are interconnected using scattering junctions. In its canonical form, the delay time of the individual bidirectional delays is 1 sample. The purpose of the scattering junctions is to re-distribute the wave energy into connected delays based on the relative acoustic impedances of the incoming and outgoing waveguides. For example, in the case of a scattering junction in the centre of a room, the impedances of the outputs of all connected delays would be equal, therefore any incoming energy is equally distributed among all their inputs. By comparison, if the scattering junction is located at a very reflective surface, then there is a mismatch in the amplitudes of the waveforms that are routed to the connected delays, sending more power in some directions than others.

Typically, digital waveguide meshes are arranged in a two dimensional configuration as is represented in Figure 1.19. One excellent introductory description of this system is Van Duyne (1993).

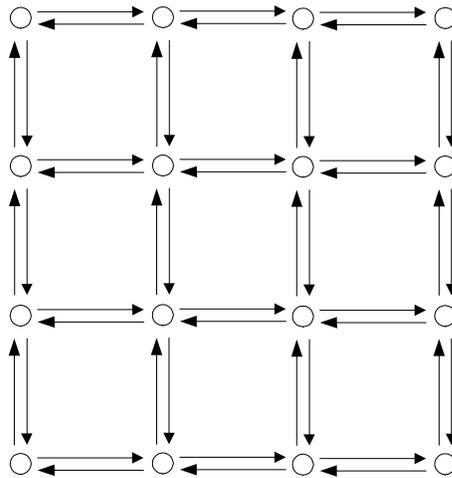


Figure 1.19 Two-dimensional digital waveguide mesh. Each circle denotes a scattering junction with four connected bidirectional delays (Van Duyne 1993).

There are a number of problems associated with the implementation of digital waveguide meshes such as dispersion and quantization error. One particular difficulty with using the discrete rather than a continuous representation of space that occurs in the mesh is frequency-dependent differences in wave propagation speed in different directions on the mesh. This is particularly noticeable in two-dimensional meshes with square layouts as shown in Figure 1.19. In such a configuration, diagonally-travelling waves have a frequency-independent propagation speed. Low-frequency waves travelling along either of the two axes will have an identical speed to their diagonal counterparts, however, high frequency information traveling along the axes has a speed of 0.707 that of lower frequencies. This results in a distortion of the wavefront in the form of a loss of transient information in some propagation directions.

One solution to this problem is to increase the number of dimensions on the mesh. Although the system is still used to model a two-dimensional surface, it contains a larger number of axes and thus reduces differences in wave speed with direction. One example of such a system is shown in Figure 1.20 which displays a triangular mesh. Savioja (1999) provides an evaluation of various mesh topologies and the resulting wave speeds.

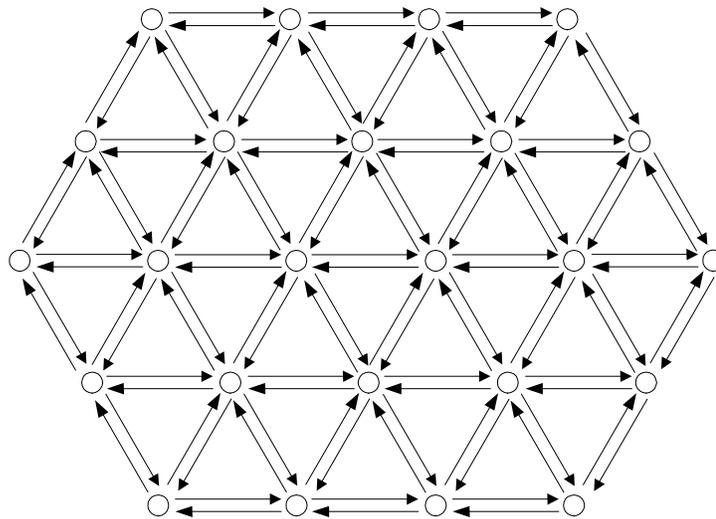


Figure 1.20 Triangular digital waveguide mesh

Digital waveguide meshes also suffer from cumulative quantization errors due to the multiplicity of parallel and series combinations of gain modification in the junctions. One suggested solution to this issue is the re-distribution of the error into the spatial domain. In this topology, the quantization error is traded for a much less significant dispersion error on the mesh (Van Duyne 1993).

Unfortunately, although physical modelling reverberation systems can produce excellent results, the processing power required to simulate the characteristics of large three-dimensional spaces is presently prohibitive in real-time systems.

1.5.4 Convolution through a measurement of real space

In 1979, Moorer (1979) discussed the option of using a measurement of an existing space as a digital reverberator. In this system, an impulse response of a real room is measured and stored, and the characteristics of the measurement system are subtracted. This impulse response can subsequently be used as an FIR filter that can be convolved with an anechoic recording to produce a result equivalent to the recording reproduced in the space. More recently, it has become possible to implement such a system in real time as is exemplified by the new DRE-S777 “sampling reverb” unit from Sony Electronics.

There are two possible methodologies for performing the actual convolution known as *real convolution* and *fast convolution*. In the first, the impulse response is effectively treated as a multitap delay with a large number of taps (equal to the length of the impulse response in samples) and a gain control on each. Although this system can provide an output with extremely low latency, it typically requires an enormous amount of processing power. Consider for example, that in the case of a 2-second reverberant tail sampled at 44.1 kHz, a processing system using real convolution working in real time on a standard floating point processor with a large cache would be required to perform approximately 7.8 billion operations per second per channel.

Fast convolution relies on the interrelated properties of the frequency and time domains. Since the complex multiplication of two signals in the frequency domain is equivalent to the convolution of the same signals in the time domain, it is possible to perform an FFT on each signal simultaneously, multiply their respective real and imaginary components, and performing an inverse FFT on the product, producing a temporal convolution of the signals. This procedure results in a considerable savings in operations (844 million operations per second for the same configuration as the previous example with an FFT size of 2048 samples), however it does incur a number of issues to consider such as latencies caused by the FFT block size and strategies for dealing with overlapping results of the IFFT process.

There are two principal disadvantages of a convolution-based reverberator:

1. **Parameter manipulation:** If the system is based on a measured impulse response, then it is impossible to alter characteristics of the reverberant space in a believable manner. For example, in the Sony device mentioned above, the user is offered a control over the reverberation time, however, this merely controls a decay envelope which is applied to the measured impulse response before convolution. While this emulates an altered absorption coefficient for the reflective surfaces in the space, it can result in some drastic effects such as the elimination of all but the early reflection patterns.

2. **Processing power:** In order to convolve signals through an impulse response in real time with a zero latency, real convolution must be used. While this is possible, it requires computational power that, until recently, has been prohibitive. This will cease to be an issue in the near future for 44.1 kHz and 48 kHz, however, it will be a number of years before higher sampling rates are feasible.

There is one overriding advantage in using this topology: sound quality. No artificial reverberation developed to date matches the realism of a system that uses correctly prepared measured impulse responses of real spaces.

1.5.5 Stochastic model

In theory, a reverberant decay approaches a perfectly diffuse field; in fact, many linear acoustics texts assume that the reverberant tail is a diffuse field. While informal listening tests and discussions in the MARLab indicate that this may be overly simplified, (Bech 2000) (Dalenbäck 2001) it can be used as an approximation. For example, in a TRD model, allpass and recursive comb filters are used to create a perceptual equivalent to a room's reverberant tail. When a device is intended for a multichannel output, manufacturers typically alter the parameters of these filters in order to achieve the required number of uncorrelated, but perceptually similar signals in an effort to simulate this theoretical diffuse field.

In 1998, Rubak and Johansen suggested that an impulse response of this field could be simulated using a recursive pseudo-random algorithm with an applied decay envelope.

According to preliminary listening tests, they reported that a pseudo-random reverberator could produce high-quality reverberation with a minimum echo density of 2000 per second and a repetition rate of no greater than 10 Hz (Rubak and Johansen 1998) (Rubak and Johansen 1999).

Informal experiments in the MARLab prove that this method does indeed produce good results, however an extension on this model is recommended. A more realistic and controllable decay curve can be achieved by dividing the uncorrelated noise samples into constituent frequency bands and applying a different exponential decay envelope to each, typically using shorter times for higher frequency bands.

One excellent characteristic of this algorithm is the automatic decorrelation of the various output channels when different original noise samples are used. This contrasts with the sometimes painstaking attention to the parameter relationships required to achieve the same isolated result with unit reverberators. It should be noted, however, that although this system provides results equalling or possibly improving on the TRD model, it is considerably more computationally expensive and therefore unattractive as a useful procedure.

1.6 Review of documented methods of diffusing early reflections

A number of researchers have developed various schemes for the simulation of diffused early reflections in synthetic acoustics systems, particularly for auralization. Most of these are included in the excellent summary by Dalenbäck (1994). In this, he divides the procedures into ten different categories based on combinatorial complexity and methodology. Sections 1.6.1 through 1.6.10 inclusive are a summary of this paper, with the section titles taken directly from the paper. Note that Dalenbäck is considering the behaviour of the various systems for higher-order reflections off combinations of specular and diffusive surfaces. As a result, some titles include this behaviour through categorizations such as “S-D” (Specular-to-Diffusive surface). The descriptions and evaluations in this dissertation will concentrate on the issue of the systems’ behaviours for first-order reflections.

There are two commonly-used characteristics in models for diffusing reflections. The first, which is used in almost all of the described procedures is the presence of stochastic functions in place of a mathematical model. In order to reduce processing requirements, many systems attempt to simulate the perceptual characteristics of diffused reflections without the detail of a physical model, with varying degrees of success.

The second is the interdependency of the decay curves of the diffused reflection and the overall reverberation. The two principal equations used for predicting reverberation times in rooms were developed by Sabine and Eyring (Kuttruff 1991) Both of these functions are based on the assumptions that all surfaces within the enclosure are perfectly diffusive

(Mehta 1976). As a result, many models assume that all diffused reflections must therefore decay at the same rate. This presents an iterative problem with predictive acoustics systems, since the model being used to determine the reverberation decay time is based on the decay time of the reflections which are, in turn, determined by the reverberation decay time and so on.

1.6.1 Ray tracing with randomized ray directions

The first, and possibly simplest method of simulating diffused reflections is to assume that the diffusing surface acts as a frequency-dependent specular reflector that re-directs the wave in a random direction (Kuttruff 1980). This system is initially implemented using a ray trace procedure, however the outgoing direction of the ray from the diffusive surface is randomized. This procedure is implemented in multiple frequency bands, typically at octave spacing (Dalenbäck 1994). In cases where a surface is partially diffusive, a random function is used to decide whether the reflected ray will be processed as a specular or diffused reflection.

This system produces good results for late components of the room impulse response, (Dalenbäck 1994) however early reflections have an essentially specular characteristic and are therefore unsuitable for simulation of diffused first reflections.

1.6.2 Ray tracing where the diffused energy decays with the reverberation time of the room

This methodology begins by using either the Sabine or Eyring equations (Kuttruff 1991) to predict the decay rate of the entire room. This value is then used to create an algorithm with a matching decay time. The results of a ray trace model are processed by this algorithm before arriving at the receiver, thus the time smearing effect of diffused reflections is simulated, however the spatial distribution of the power is neglected.

This method is used by a number of people (Mehta 1976) (van Rietschote 1981) however, in all cases, the specific method of creating the algorithm for diffusion is not described. However, one system description (Nakagawa 1993) notes a commercially available hardware reverberation unit routed to a subset of loudspeakers in a custom multichannel configuration to produce the diffused reflection signal.

1.6.3 Ray / cone tracing where the diffused part of the energy initiates a D-D reflection process

This method, developed by Malcurt and Jullien (1986) and refined by Warusfel et al. (1992), is intended for use in acoustics prediction systems. In the first system, the reflective surfaces are subdivided into discrete patches which are used as reference locations for a cone tracing system. Two “passes” are made, the first to determine the behaviour of the specular component of the reflection, the second for the diffused component. In the first pass, the specular component is reflected to the receiver or to a

1.6.3 Ray / cone tracing where the diffused part of the energy initiates a D-D reflection process

second surface, according to the geometry of the space, and the diffuse component is “stored” at the surface for the second pass. The diffuse radiation of the surfaces is propagated using a Markov process using a discretized version of Kuttruff’s integral equation (Kuttruff 1991) and assuming a perfect Lambertian behaviour (Malcurt 1986). The Markov process, when implemented in discrete time, is one where the probability of the state of the system at a given time $t+1$ is dependent only on the state of the system at the previous time t (Sollich 2001) (Rabiner 1986). This results in a stochastic model whose output corresponds to the statistical behaviour of a diffused reflection.

The refinement of the model by Warusfel et al. (1992) improves only on the accuracy of the specular reflection component by implementing an image model instead of the cone trace method.

Note that the “D-D reflection process” in the title refers to the fact that both of these models assume that, after the second order reflection, all surfaces are perfectly diffusive.

These models result in a good spatial distribution of energy for the diffused reflection and are not reliant on the Sabine or Eyring equations, and can thus be used for the prediction of reverberation times. However, they provide a stochastic model with appropriate probability of energy distribution and are thus useful only on a time scale on the order of the reverberation rather than of the initial reflections.

1.6.4 Image source clouds

If we assume that a diffused reflection can be considered as a series of spatially distributed specular reflections as is shown in Figure 1.7, then the image model can be used to simulate the resulting multiple reflections. In effect, the model considers a “cloud” of sources at locations determined by the relationships between the sound source, reflective surface and receiver (Dalenbäck 1992) (Lehnert 1992) as is shown in Figure 1.21. As such, the model results in a smearing of the impulse response of a reflection, however a number of issues arise. Firstly, the Lambertian behaviour of a diffusive surface is typically not included in the model since the diffuse reflection is modelled as a cluster of specular reflections. In addition, a problem for predictive acoustics systems is that the simple issue of determining the amplitude of the individual reflections is unclear in this model.

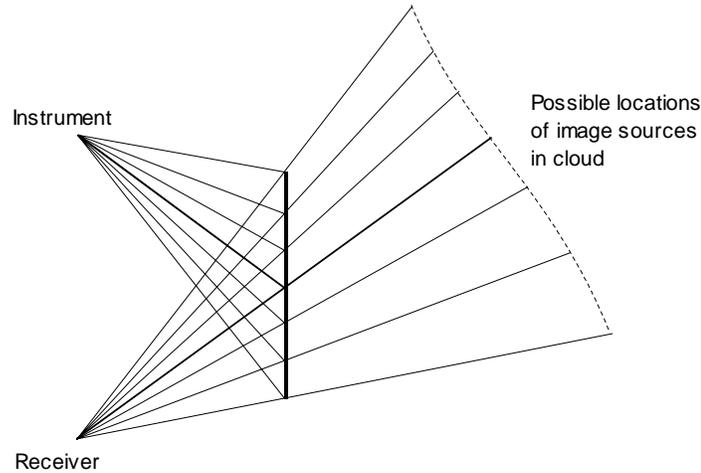


Figure 1.21 Image source cloud method. The simulated diffused reflection consists of a cluster of selected images located along the dotted line on the right of the diagram.

1.6.5 Filters on specular paths

This method is a hybrid of the methods described in Sections 1.6.2 and 1.6.4 (Dalenbäck 1994). In it, a specular reflection is first calculated using the image method. This impulse response is subsequently processed by a filter that produces a diffused cluster before sending the signal on to the receiver (Borish 1984). In effect, this is similar to processing each specular reflection with a one-channel TRD-based reverberation device, resulting in temporal smearing. The relative simplicity of this implementation is the reason for its use in commercial reverberation devices as will be discussed in Section 1.7.1.2. Since the “diffusion” is produced using a single channel there is no spatial distribution of the signal: the entire cluster is located at the point of specular reflection. This system has a number of drawbacks and a single principal advantage. Informal listening tests in the

MARLab indicate that the spatial distribution of diffused reflections is of primary importance, consequently this system is not adequate. It is, however, widely used, particularly in real-time reverberation systems due to its relatively low computational demands.

1.6.6 Hybrid method combined with separate ray tracing

This method separates the reflections according to their specular and diffuse behaviour and calculates each independently. The specular reflections are determined using the hybrid method while the diffuse reflections are computed using ray tracing with randomized directions (Rindel 1997) as described in Section 1.6.1, and thus suffer from the same problems as described earlier.

1.6.7 Early part, hybrid method / late part, modified ray tracing

This method divides the early and late parts of the impulse response using a “transition order” (Dalenbäck 1994), and treats them separately. For reflections below the transition order, the reflections are considered to exhibit a specular behaviour and are calculated using the hybrid model. The late part of the response is constructed using a method based on ray tracing with randomized directions (Naylor 1993). In this case, however, the direction of the outgoing wave is not simply randomized, rather the point of reflection is checked to determine whether it is “in sight” of the receiver. If it is, then the ray is traced directly to the receiver using a perfect Lambertian model as described in Equation 1.17 without incurring a loss of power in the propagating wave. The advantage of this method

is that a high number of reflections are produced without the expense of tracing a large number of rays, however, the method suffers from the same problems as the simpler system described in Section 1.6.1.

1.6.8 Wall subdivision using specular and diffusing nodes

This method begins by subdividing the reflective surface into interconnected “nodes,” each with specular and diffusive properties (Lewers 1993). Each of these nodes is considered to be connected to the sound source and receiver nodes similar to a ray trace model with discrete surface locations rather than angles of radiation. The power from the sound source node is distributed to each surface node which, if it is completely specular, re-directs the wave to the receiver node accordingly. However, if the node has diffusive properties, the wave is randomly redistributed to other surface nodes in addition to the receiver node. One drawback of this system is that a very high number of surface nodes is required in order to achieve an adequately high density of diffusion for low-order reflections. Furthermore, very low-order reflections result in a highly specular characteristic.

1.6.9 Early part image model plus S-D combinations, late part ray tracing

In this model, the early reflections are produced using an image model while the late components are calculated using the ray tracing technique. These late reflections are diffused by randomizing the directions of the reflections as was described in Section 1.6.1. The difference in this model lies in its method of synthesizing first-order diffused

reflections and second-order specular-to-diffused combinations. For these two situations, the surfaces themselves are considered to be comprised of radiating sound sources, a technique described by both Dalenbäck (1992) and Rindel (1997). The latter contains a summarized description as follows: “the rays are treated as transporters of energy rather than explorers of the geometry.” The outgoing rays from the sound source meet the reflecting surface which is then considered to be a new source at the collision point. The subsequent behaviour of the source is dependent on the model. One possible method which is rarely implemented due to computational requirements is to convert the local impedance response of the surface at the collision point into an impulse response corresponding to the temporal characteristics of the radiator. The collision point can then be considered to be a filter through which the incoming signal is processed and re-radiated towards the receiver. This is the method used in the model proposed in Chapter 3. A second possible behaviour is described below.

1.6.10 Split-up of rays reflected on diffusers

This is a method of diffusion using ray tracing in a system with multiple frequency bands (Dalenbäck 1996) and can be considered a substantial refinement of the system described by Lewers (1993). Specular reflections are calculated using a system described as *approximate cone tracing*. The “approximate” qualifier is added because the cone face calculations are based on the characteristics of the reflecting surface at the centre of the cone rather than the entire area. The diffused reflection is calculated differently. The diffusively reflecting surface is assigned a series of absorption and diffusion coefficients

(α and δ respectively) in octave bands. It is then subdivided into square patches, each of which exhibits the characteristics of the surface and acts as an independent secondary sound source.

The system is based on a series of “passes” along the reflective surfaces in the room. In the first pass, the power incident on each patch directly from the sound source and from specular reflections is stored and multiplied $\delta(1-\alpha)$ of its original values in each frequency band at the appropriate times in the local impulse response. This response for the various incident rays is labeled the “primary list” (Dalenbäck 1996). Once the entire first pass is completed, the centre of each patch is considered as a discrete sound source radiating its impulse response into the space on a number of rays. Each of these rays is traced either to the receiver (where the entire primary list is added to the overall impulse response), or to another reflective surface (where it is treated in the same manner as the incident power in the first pass). In the latter case, the entire diffused response from the first patch is attenuated and re-radiated into the room after the second pass. This process is repeated until a threshold determined by the user is reached.

This process results in a spatial distribution of the diffused reflection power as well as control over the frequency-dependent diffusion characteristics of various surfaces. In addition, it permits the user to assign different diffusive characteristics to different locations on a reflective surface, as is usually the case in modern concert hall designs. One drawback of the system, however, is that it does not accurately model the local impulse response for each patch on the surface, opting instead to provide a general

statistical model. This is excellent for higher-order reflections and reverberation tails, however, it is not representative of the microscopic characteristics of the low-order diffused reflections.

1.6.11 Digital waveguide mesh with modulating ringguides

The use of a digital waveguide mesh as described in Section 1.5.3.3, either in two or three dimensions, has the restriction that the boundary geometry must match the lattice structure of the mesh. When the simulated reflective surface cannot be drawn along the bidirectional delay lines, *ringguides* must be added to extend the perimeter nodes to the closest point on the boundary as shown in Figure 1.22 (Laird 1998). Unlike the mesh, the ringguides are non-integer length waveguides comprised of an integer delay and a fractional delay modelled using a first-order allpass filter.

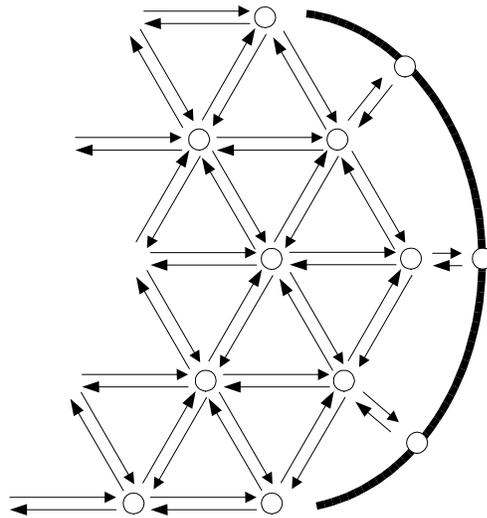


Figure 1.22 Two dimensional triangular digital waveguide mesh with rimguides extending to junctions on the curved boundary.

It has been proposed that the effects of diffused early reflections can be simulated by modulating the rotation of the rimguides randomly on each sample step (Laird 1999). The amount of diffusion is then determined by the amplitude of the modulation. One problem associated with this method is the presence of noise generated by large amounts of modulation (Laird 1999), however it does have the possible benefit of providing the user with a time-variable (and potentially frequency-dependent) diffusion on a digital waveguide mesh.

1.6.12 Convolution through measured impulse responses

One method based on the system described in Section 1.5.4 is to convolve a simulated reflection through a measured impulse response of a reflection off a real diffusive surface (Dalenbäck 2001). While this method can provide good results, it suffers from the

inability to modify parameters to any characteristics which do not directly correspond to previously measured geometries.

1.7 Typical implementation of early reflection patterns

The ever increasing power of DSP engines available to the consumer has meant the inclusion of more and more sophisticated algorithms for synthesizing room acoustics in recording studio environments (Lund 2000). One of these developments has been the widespread inclusion of simulated early reflection patterns in relatively low-cost processors.

Current manufacturers typically opt for one of the three following basic topologies for the implementation of these simulated early reflection patterns in commercial units.

1. Provide the user with a group of delays with factory-preset gain and time relationships, typically based on ray-trace models. While the selection of the particular reflection set and its global gain relative to the reverberant tail is usually adjustable by the user, there is rarely an option to access individual reflections within the set. These collections of delays are normally based on a room model which has been chosen by the manufacturer and are selected by generic room type with names such as “Hall,” “Studio,” or “Church.” This system presents advantages both to the manufacturer and to the consumer in that it reduces the complexity of the user-controlled functions for the reflections. However, the manufacturer must pay particular attention to the specific temporal relationship between the individual reflections in order to avoid undesirable resonances, ringing or audible comb filtering effects.

2. Provide the user with a number of independent delay lines, each with adjustable delay time, filtering and localization (panning). This system provides the user with the most detailed control over the early reflections and their relationships both to each other and to the reverberant tail. As a result, however, it also requires the most attention on the part of the end user, and is likely to be less attractive in a typical busy production studio environment. Assuming that the processor has adequate power and memory available to make such a system feasible, this system requires the least attention on the part of the manufacturer, who needs only to provide the processing to the user without making any pre-determined aesthetic decisions regarding the preset. This implementation is typically found only on higher-end, and therefore higher-priced reverberation units, and is usually implemented in addition to the first method.

3. Provide the user with control over the dimensions and surface materials of the virtual room which is being simulated, and calculate the appropriate location, time and gain relationships accordingly. This requires little knowledge on the part of the users regarding the particulars of signal processing, freeing them from the necessity of thinking in a time domain, rather presenting a more intuitive, familiar interface relating to real acoustic spaces. Moreover, the aesthetic tailoring done by the manufacturers described in the first procedure can be relegated to the processor which can be programmed to determine in real time the appropriate delays to use for the synthetic model. Although this method was introduced in

1992 (Gardner 1992b) it has not been widely implemented due to computational complexity, particularly for real-time applications. As processing power increases, however, this is a system which will likely become the archetype. It is presently offered commercially only on software “plugins” for digital audio workstations. Note that this is also the interface for the second and third incarnations of the MARLab SceneBuilder software.

In all of these implementations, due in large part to restraints on computational power, most manufacturers use simple multitap delay lines to produce their reflection set. Typically, the panning of these delay taps is distributed across the full width of the sound stage to ensure that there is a minimum interchannel cross correlation, and therefore a maximum perceived spaciousness for the algorithm (Kurozumi 1983). The localization is typically accomplished using simple power panning to ensure compatibility with monophonic playback. The individual delay times are chosen to avoid simple relationships between them in order to reduce the number and audibility of the resulting interference resonances.

In some cases, a modicum of filtering is performed on the delays to simulate the absorptive characteristics of the reflecting surfaces. This is usually limited to simple high and low pass filters which are applied to the input to the multitap in the interest of economy, resulting in multiple identical copies, as will be seen in the following section.

1.7.1 Impulse response analyses

There is a dichotomy between the traditional and current methods of producing early reflection patterns in DSP-based reverberation engines and our understanding of real acoustics. For the past 100 years, it has been well-understood that there is a correlation between diffused reflections and better-sounding concert halls. By comparison, since the introduction of early reflection patterns to synthetic reverberation algorithms by Schroeder in 1970, processing devices have been programmed to simulate early reflections with the assumption that the reflecting surfaces in the synthetic environment are almost perfect specular reflectors.

In order to demonstrate the requirement for a new system for generating early reflection patterns in synthetic acoustic environments, the response of two typical professional stereo reverberation units are analyzed. These have been chosen as devices representative of those found in many recording and post-production studios and, while some characteristics of the units are particular to the manufacturers, the general characteristics are ubiquitous across almost all brand names.

1.7.1.1 t.c. electronic M3000

Figure 1.23 shows the impulse response of a standard reverberation patch on a t.c. electronic M3000 measured using a custom impulse response recorder constructed in Max / MSP (Parameter settings are listed in Appendix A). The early reflection pattern has been intentionally separated from the dense reverberant tail by the application of a 200

msec predelay to the latter merely in order to more easily delineate the two visually. Note that, without the excessive predelay, the later delays in the early reflection cluster would occur concurrently with the beginnings of the reverberant tail at roughly the same amplitude.

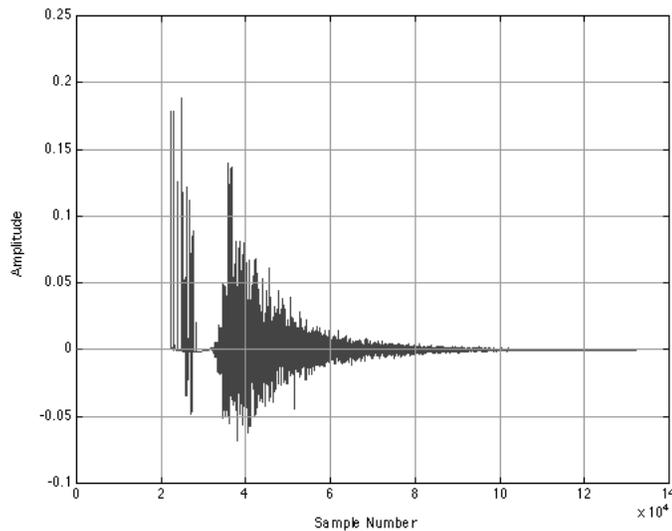


Figure 1.23 Amplitude vs. Time plot of the impulse response of a reverberation patch on a t.c. electronic M3000 stereo reverberation processor ($F_s = 44.1$ kHz).

Figure 1.24 displays a detail of the impulse response, showing only the isolated early reflection pattern before the onset of the reverberant tail. There are a number of characteristics of this delay cluster which should be discussed. To begin, the relative time relationships of the 33 individual delays exhibit no obvious simple relationships, thus avoiding audible resonances. This is confirmed in a frequency response analysis. This leads to two conclusions. Firstly, it is reasonable to conclude that recursive delays are not

used to create the cluster, rather individual parallel delays or, more likely, a single multitap with independent output gains. Secondly, it is obvious that the manufacturer has taken some care in choosing the relative times of the delays. Were this not the case, the more computationally economic route of recursion would have been taken.

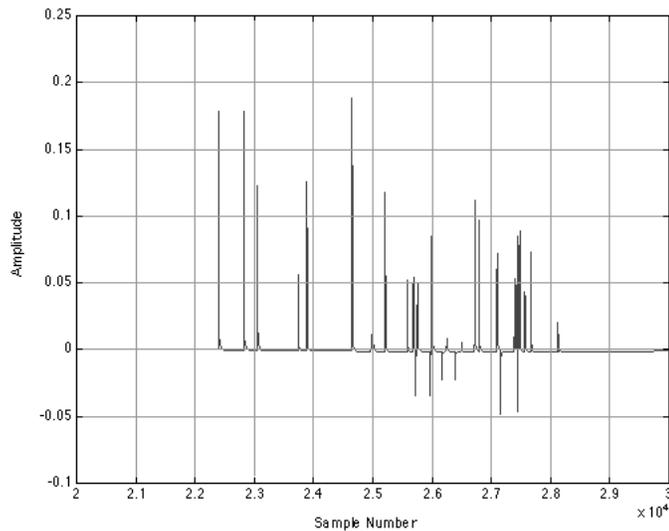


Figure 1.24 Detail of the previous plot, showing the isolated early reflection pattern.

It is also curious to note that the individual envelopes of each delay tap, an example of which is shown below in Figures 1.25 and 1.26, are identical. This leads to the conclusion that a single filter has been applied to the signal likely somewhere in the processing chain before the input stage of the delay(s), resulting in a series of multiple independent taps exhibiting identical frequency characteristics.

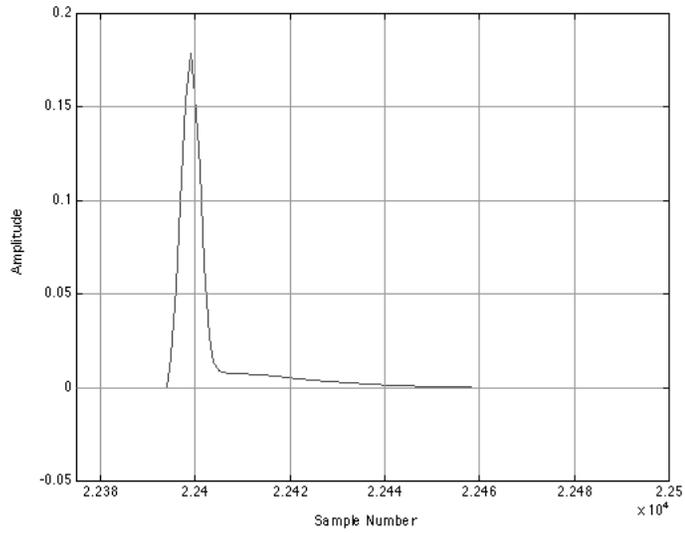


Figure 1.25 Detail of the previous plot, showing isolated first reflection. All individual reflections in the early reflection cluster use this envelope.

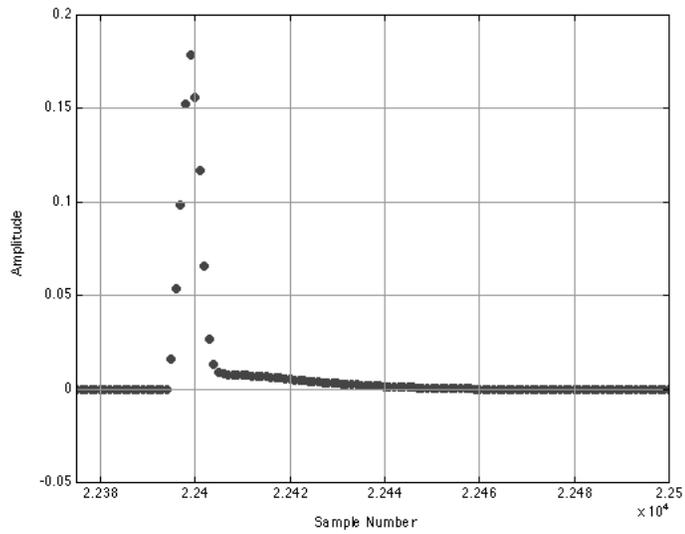


Figure 1.26 Plot of data identical to previous plot showing amplitudes of individual samples.

To derive the characteristics of the filter used to modify the input signal, we can use a Fast Fourier Transform (FFT) analysis to process the isolated single delay output shown in Figure 1.26 in MATLAB. In order to do this, the data comprising the isolated reflection were placed in the centre of a vector otherwise consisting of zeros. Using a 65,536-point FFT, the signal was converted to the frequency domain, from which was derived the power spectral density (PSD) (MathWorks 1998) of the signal which is plotted in Figure 1.27. This analysis was subsequently verified in the analog domain using a DRA Laboratories MLSSA.

There was a small time variant negative DC offset on the signal, the result of a high pass filter and the predominance of positive values in the impulse response of the early reflection cluster. This offset was removed since it can not be accurately analyzed using an isolated single reflection due to its level dependence on adjacent reflections. The only effect of this modification is the apparent extension of the low frequency response of the signal. In addition, an arbitrary gain was applied to the signal to raise its level to an appropriate level for plotting. This gain has no effect on the relative power levels of the plotted frequencies.

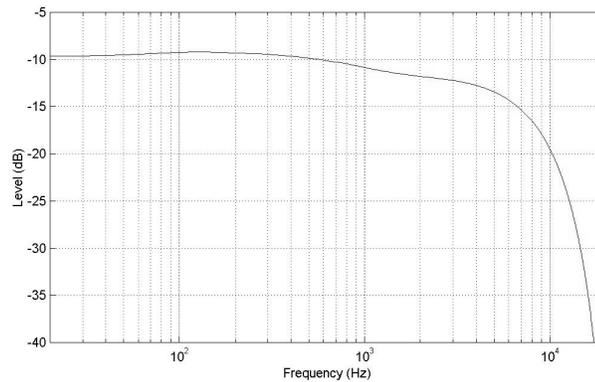


Figure 1.27 PSD plot of the single reflection plotted in Figure 1.26.

There are two conclusions to be drawn from this analysis. Firstly, since all reflections in the pattern have the same envelope and therefore frequency response, we can conclude that the device is not intended to mimic the cumulative frequency-dependent absorption effects of higher-order reflections. In effect, all of the 33 taps in the cluster are considered in this regard to be first-order reflections. Secondly, the minimal effect of the filter on the frequency response of the early reflection generator indicates that the simulated reflection coefficients are of a highly reflective surface.

1.7.1.2 Lexicon 480L

Unlike almost all real-time reverberation devices, the Lexicon 480L offers an option for the user to adjust a level of “diffusion” applied to the early reflection patterns in reverberation programs. This process is described in the users’ manual as a “cluster of preechoes” (Lexicon 1993) whose density is varied using the DIFFUSION controller. This density parameter is controlled in the later part of the reverberant decay using the SIZE option, however, only the early reflections will be examined for this dissertation.

An impulse response measurement of a 480L reverberation program (Parameter settings are listed in Appendix A) reveals that there are a number of early reflections incorporated as part of the standard reverberation algorithm, however, the user is offered an option to include two or six additional delays (depending on the particular program used) with adjustable times and levels. Since the diffusion process on a given output channel is the same for both the preset and user-defined early reflections, the latter was used for these measurements. This choice was made in order to simplify an examination of the relationship between the two main output channels.

The temporal response of the early reflection delay tap is shown in Figure 1.28 and the corresponding frequency response in Figure 1.29.

1.7.1.2 Lexicon 480L

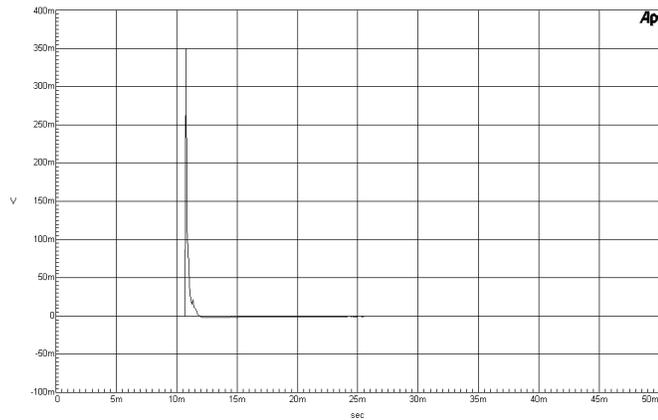


Figure 1.28 Impulse response of a single early reflection in a Lexicon 480L. Left channel, delay time = 8 ms, level = full, diffusion = 0.

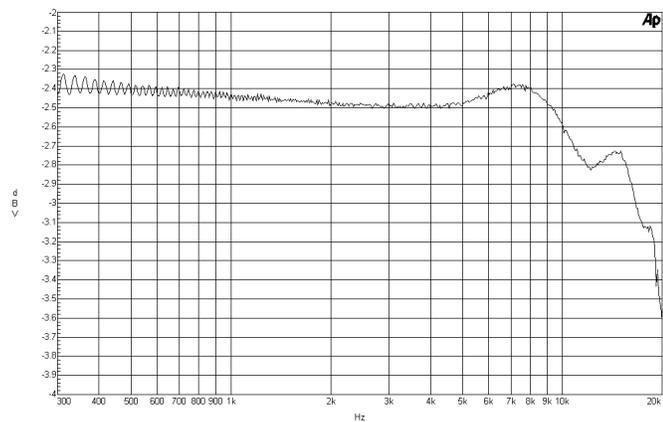


Figure 1.29 Frequency response plot of isolated early reflection plotted in Figure 1.28.

The increase in level of the diffusion parameter reveals that the control is a simple constant power level adjustment between the outputs of the simple delay tap shown above and of the diffusion processor. This is evident since, as the level of diffusion is increased, the level of the specular reflection decreases. This can be seen in a comparison of the graphs in Figures 1.28, 1.30 and 1.31.

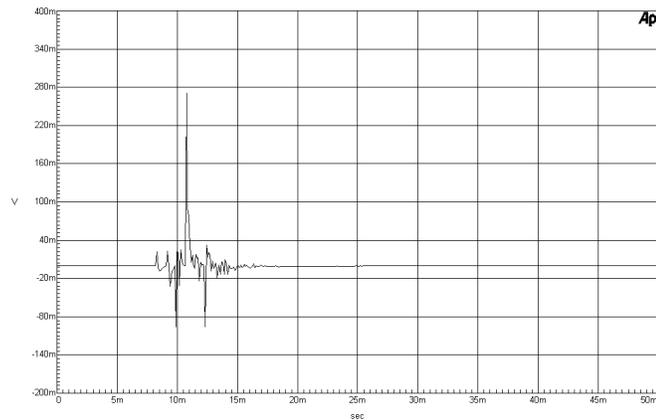


Figure 1.30 Impulse response of a single early reflection in a Lexicon 480L. Left channel, delay time = 8 ms, level = full, diffusion = 50. Note the drop in level of the reflection compared to Figure 1.28 as well as the beginning of the diffuse cluster before the specular reflection.

One characteristic of the diffusion of the early reflection in the 480L which raises suspicion is the time of arrival of the onset of the impulse response. As can be seen in Figure 1.30 and in comparisons between Figures 1.28 and 1.31, the diffused reflection begins *earlier* than the specular component. As was discussed in Section 1.2.1.2, this is not the case in real-world situations. The spatial and temporal distributions of power in a diffused reflection will vary according to the particular surface characteristics and relative locations of the source, surface and receiver. However, the specular reflection will always originate from the location on the surface with the shortest propagation distance of any reflection and will thus arrive at the receiver earliest. Consequently, we can begin to surmise that the diffusion model implemented in the 480L is based on a perceptual rather than a physical model.

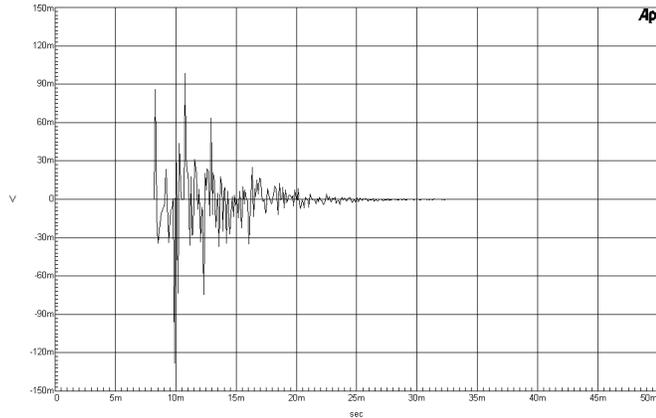


Figure 1.31 Impulse response of a single early reflection in a Lexicon 480L. Left channel, delay time = 8 ms, level = full, diffusion = 99.

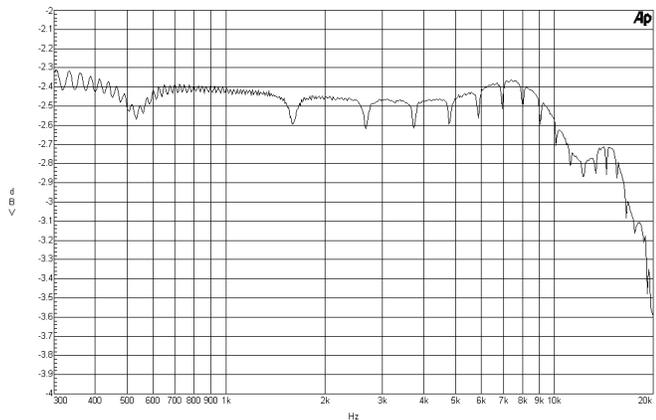


Figure 1.32 Frequency response plot of isolated early reflection plotted in Figure 1.31. Note the small effect the high level of diffusion has on the frequency response relative to that shown in Figure 1.29.

Figure 1.32 shows a frequency response plot of the isolated fully diffused reflection. Note that there are only very small deviations in this response relative to the fully specular reflection, indicating that the procedure used to create the model is tailored to have little

to no effect on the perceived timbre of the delay. With the exception of an added periodic notch of approximately -0.15 dB at intervals of approximately 1 kHz and an inversion of the low-level ripple, the frequency responses match almost exactly, partially indicating a use of allpass filters to produce the effect. This suspicion is confirmed by an inspection of the temporal response of the diffusion component. The impulse response of the diffuse cluster appears to be based on a number of recursive delay components whose feedback delay times are determined by the `SIZE` parameter. As can be seen in Figures 1.33 through 1.35, the changing delay times with varying size values is evident. Although complete deconstruction of the cluster is impossible, the algorithm appears to be comprised of approximately two to three allpass filters and as many recursive or non-recursive comb filters.

1.7.1.2 Lexicon 480L

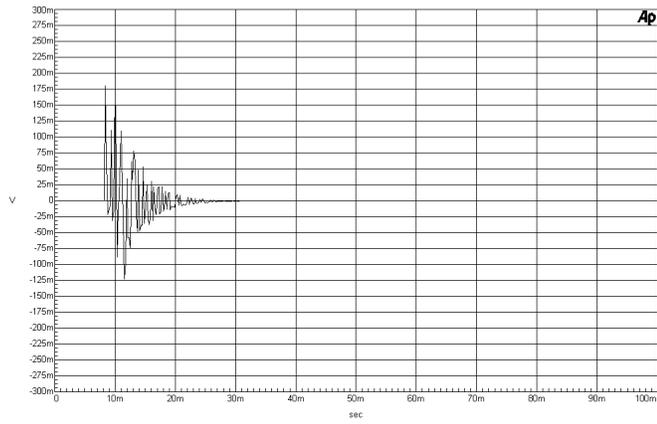


Figure 1.33 Diffusion component for an 8 msec pre-echo, size = 4 m

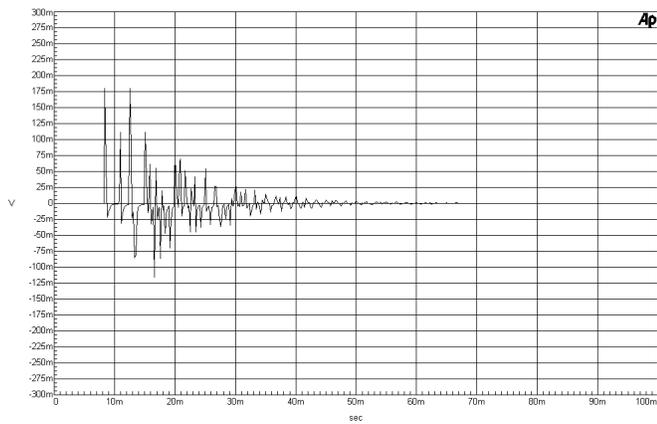


Figure 1.34 Diffusion component for an 8 msec pre-echo, size = 10 m

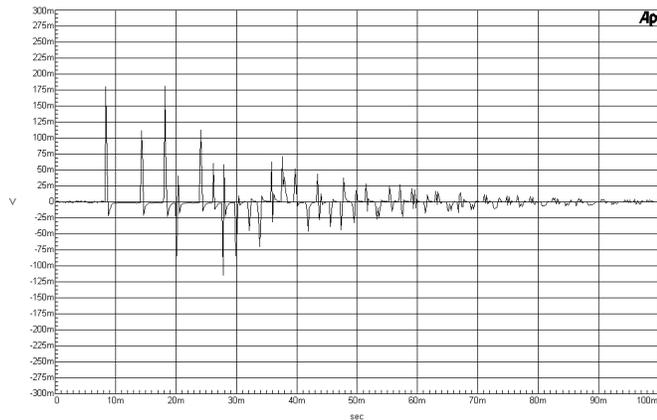


Figure 1.35 Diffusion component for an 8 msec pre-echo, size = 25 m

Also worthy of note is the similarity of the diffusion effect throughout the impulse response. The characteristic envelope of the diffusion is identical for a given output channel regardless of the location of the early reflection in the time domain as is demonstrated in Figure 1.36. This envelope is therefore produced by a filter located in the input stages of the early reflection processor. This is evident since the same envelope is applied to all delays in the left output channel regardless of the input channel. In addition, it should be noted that the effect is only present on the output channel corresponding to that of the specular reflection. In other words, when a specular reflection in the left output is diffused, none of the diffusion components are routed to the right channel of the device, resulting in a smearing of the temporal response without the corresponding spatial distribution of the diffused reflection. These two characteristics of the system indicate that the method used to produce the effect is a filter on a specular path as described in Section 1.6.5.

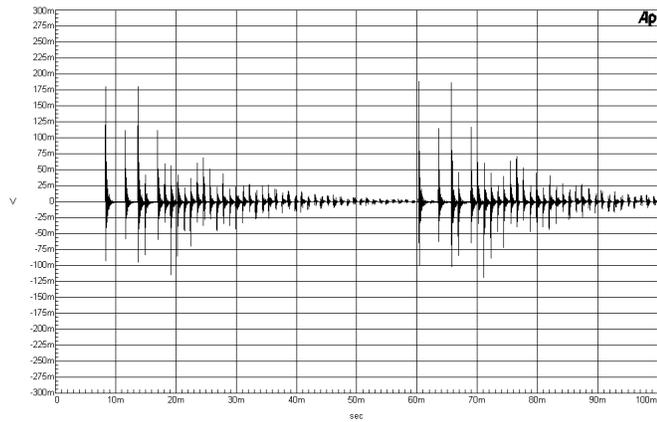


Figure 1.36 Comparison between diffusion applied to two different delay times for the left channel output. Note that the details of the envelopes are identical, indicating a single filter being applied to both delays (Diffusion = 99, Size = 14, Delays full on at 8 msec and 60 msec).

1.7.2 General assumptions

The analysis of the early reflection pattern in the M3000 demonstrates a number of characteristics which are typical of almost all synthetic reverberation engines. Firstly, it is evident from the envelopes of the individual delay taps that all reflecting surfaces modelled by the manufacturer are excellent specular reflectors. The computational economy of a simple delay line dictates that it be used in preference to the simulation of a more diffused reflection which would require both more DSP power and programming time. This justification makes an interesting parallel to the situation in modern construction of real spaces where economical concerns have resulted in large, flat surfaces and therefore specular reflectors for walls. Regrettably, although architects and acousticians have since realized that the value of diffusely reflecting walls justifies their cost, it is evident that the programmers of real-time virtual acoustic environments have not.

Secondly, the identical time and therefore frequency responses of the delay taps result in a simulation of reflective surfaces with identically matched absorption characteristics with no reflections higher than those of the first order. In a real environment, higher-order reflections would exhibit a changing response over time, due both to the frequency-dependent absorptive coefficients of the reflecting surfaces and to the air in the enclosure. With the simple exceptions of the individual gains and delay times of the taps, there is no attempt to simulate the change in the response of the various reflections over time.

While the 480L is a slightly more sophisticated (yet considerably more expensive...) device in that it provides the user with at least a perceptual model of diffused reflections, there are a number of limitations of the system that make it inadequate for music production. The principal problem with this implementation is the lack of spatial distribution of the diffusion, an effect which, as will be discussed later in respect to the proposed model, has proven to be of almost paramount importance. The second principal limitation is the lack of independent control of the decay rate of the diffused reflection, thereby reducing the potential to modify the perceived characteristics of the reflecting surfaces in the early part of the impulse response.

2 Three-dimensional graphics models

The effort to create believable synthetic acoustic environments parallels that for modelling three-dimensional computer graphics. Developers in both fields have been attempting to perfect computer-generated virtual environments for over thirty years, however, the focus of efforts has differed between the two fields due to the peculiarities of each. In the case of graphic models, the system is mapping the distribution of light in space. Each pixel on the computer screen is calculated to have a particular value for each of the red, green and blue (RGB) components according to various parameters such as the location of the camera, the locations, colour and reflective qualities of the surfaces in the environment, and the location and intensity of the light sources. By comparison, acoustic models are concerned with the distribution of sound not only over space, but over time as well, a consideration unnecessary in graphic modelling due to the perceptually infinite speed of light.

2.1 Models of reflection

The reflection of light off a given surface can be considered using a combination of two basic models, the *specular model* and the *diffuse model*, previously discussed in Sections 1.2.1.1 and 1.2.1.2.



Figure 2.1 Photograph of a wall surface which exhibits both specular and diffusive reflective properties. Note that there are two bright spots. The small area on the right is the specular reflection, the larger area in the centre is the diffuse component.

2.2 Shading algorithms

As computers increased in speed over the past thirty years, various shading algorithms for three-dimensional imaging were developed according to the practical and potential capabilities of the systems at the time. In order to clarify concepts which will be presented in the following chapter, a number of these shading algorithms are discussed.

2.2.1 Gourad shading model

Before 1971, three-dimensional computer graphic objects were constructed of two-dimensional polygons. The shading of each of these individual components was typically calculated according to the relationship of the angle of incidence and the surface normal using the Lambert's Law stated in Equation 1.17. To reduce the number of required calculations, light sources were typically assumed to be an infinite distance from the object being modeled. This permitted the creation of believable three-dimensional topographies in spite of the lack of smooth curved surfaces. As is shown in Figure 2.1a, the individual polygons are easily visible due to the sharp edges at their individual boundaries.

In 1971, Henri Gourad demonstrated a procedure for computing shaded pictures of curved surfaces that computed continuous gradations using linear interpolations of the intensities of the vertices of each polygon (Gourad 1971). This system was attractive because it required computation time comparable to previous systems, but resulted in apparently smooth surfaces. The images in Figures 2.1a and 2.1b, from Gourad's original paper (Gourad 1971), illustrate the system now known as *Gourad shading*. The image in Figure 2.1a required 2 minutes to compute, but the image in Figure 2.1b, smoothed using the Gourad algorithm, took only 2 minutes and 15 seconds to produce. While there are a number of problems with this shading system including a "flattening" of images, a somewhat dull appearance, and apparent Mach bands – areas of exaggerated changes of intensity (Gourad 1971) – it continues to be the shading algorithm of choice for real time

2.2.1 Gourad shading model

systems, particularly computer-based video games, due to its inherent simplicity and resulting speed.

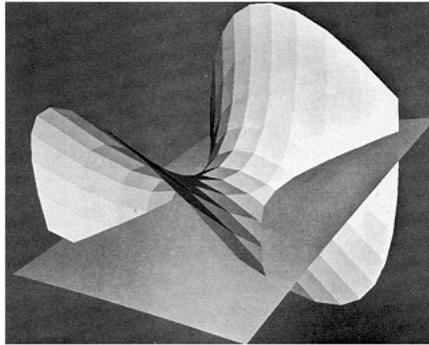


Figure 2.1a Example of Watkin's shading technique presented in Gourad's original paper (Gourad 1971).

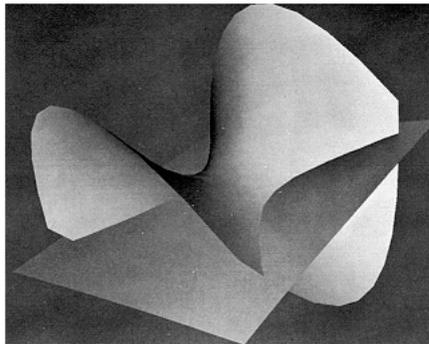


Figure 2.1b Example of Gourad's shading technique presented in Gourad's original paper (Gourad 1971).

2.2.2 Phong shading model

One of the principal problems with the Gourad shading described above is the apparent lack of depth of the images produced. This is due primarily to the fact that the intensity of

2.2.2 Phong shading model

a given pixel on the surface is determined only by the angle of the incident ray of light, thus assuming the surface to be perfectly diffuse and ignoring any contribution by a specular reflection or by the ambient light in the environment diffusely reflected off other surfaces. In 1975, Phong Bui-Toung proposed a shading algorithm that incorporated the diffuse, specular and ambient components according to a relatively simple formula (Phong 1975) (Watt 1989).

Total intensity = ambient intensity + diffuse intensity + specular intensity

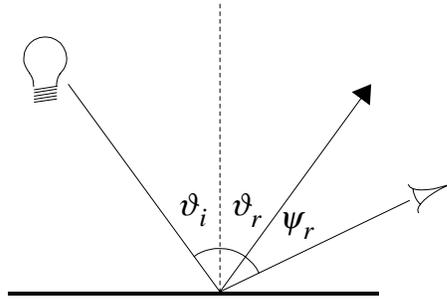


Figure 2.2 Diagram showing labelling of angles of incidence, reflectance and viewing angle.

As shown in Figure 2.2, if we label the angle of incidence ϑ_i and the angle between the viewer and the angle of reflection ψ_r , then the light intensity at the viewing position is calculated as follows: (Watt 1989)

$$I_{tot} = I_a k_a + I_i \left[k_d \cos(\vartheta_i) + k_s \cos^s(\psi_r) \right] \quad (2.2)$$

where I_{tot} is the total calculated light intensity for a point on the surface, I_a is the ambient light intensity, k_a is the ambient radiance scalar, I_i is the intensity of the light source, k_d is

2.2.2 Phong shading model

the diffuse radiance scalar, k_s is the specular reflection scalar, s is the index of smoothness of the surface and where $k_a + k_d + k_s = 1$.

As can be seen in the formula, the three constituent components of the radiant and reflected light intensities are simply added. Since the diffuse and specular components of this function assume that we are modelling a first-order reflection, the model requires an ambient component to simulate the diffuse light incident upon the entire surface due to higher-order diffused reflections from other surfaces in the modeled environment. This is a constant value added to the intensities of all points on the surface of the object. The diffuse component assumes a perfectly diffusive surface and therefore follows Lambert's Law, generating an intensity dependent on the cosine of the angle of incidence. The specular component uses an arbitrary cosine function raised to an exponent s that determines the apparent smoothness of the surface. The higher the value of s the more specular the reflection.

It is important to emphasize that this formula does not attempt to provide an accurate physical model of reality, rather it is an empirical model that produces acceptably realistic results with a combination of what are effectively ad hoc functions. Unfortunately, in spite of its relative simplicity when compared to more detailed physical models, Phong shading requires up to 100 times the computation time of Gourad shading. It does however, offer adequate quality for a considerable saving in computation time over more accurate models (Watt 1989).

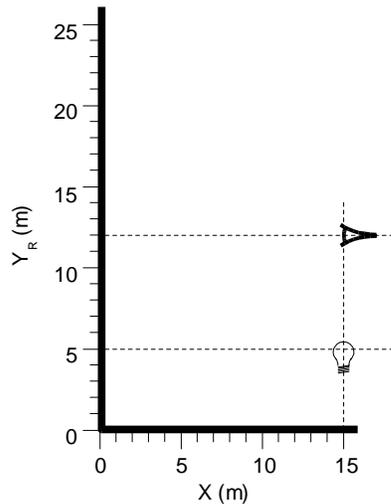


Figure 2.3 Location of light source (15, 5) and viewer (15, 12) for all of the following reflectance plots.

Shown in Figure 2.3 is a diagram illustrating the relative locations of a light source and viewer (or camera) for all plots of reflection intensities that follow. The reflecting surface is the Y-axis with reflection points in two-dimensional graphs at various values of Y_R . Unless otherwise specified, the location of the light source is 15 m from the reflecting surface and 5 m along its length at (15,5) and the location of the viewer is at (15,12). These are arbitrary locations, chosen only to avoid symmetry with the centre line of the reflecting surface.

As has been discussed, Lambert's Law relates the intensity of the reflection from a point on the surface to the cosine of its angle of incidence. Figure 2.4 shows the intensity of these individual reflections plotted as brightness for all points on the reflecting surface with an arbitrary height of 13 m and the light source and camera located at 8 m off the

2.2.2 Phong shading model

“floor.” (It may be of assistance to note the relationship between the two- and three-dimensional representation in Figures 2.4 and 2.10.) As can be seen, the brightest point in the plot, and therefore the reflection with the greatest intensity is at location $(0, 5, 8)$, the point where the angle of incidence is 0° . The intensity pattern forms a series of concentric circles around this point since the angles of incidence of all points on the same circle will be equal.

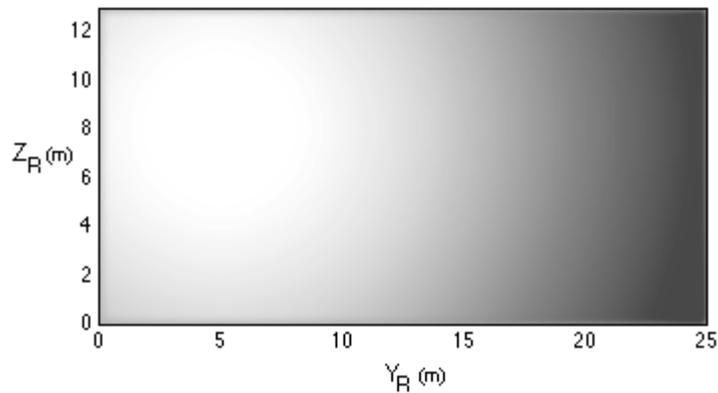


Figure 2.4 Diffused reflection of a light source at coordinates specified in Figure 2.4 and a height of 8.

If we take a one-dimensional “slice” of the plot in Figure 2.4 along the line $Z_R = 8$, we can plot a curve representing the intensity of the reflection for various values of Y_R in the two dimensional environment. This graph of intensity vs. Y_R is shown in Figure 2.5.

2.2.2 Phong shading model

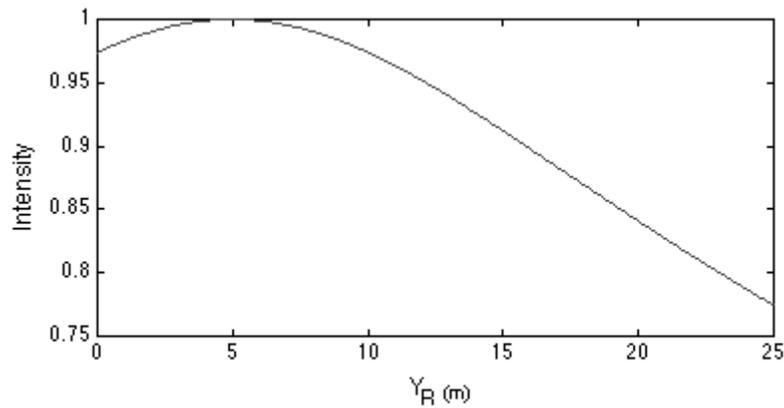


Figure 2.5 Plot of Intensity vs. location on specular reflecting wall model at $Z_R = 8$ m for locations shown in Figure 2.4.

Calculations for the specular reflection from a surface of the same size are plotted in Figure 2.6. There are two principal points to note in this graph – the location and the area of the bright spot. In the case of specular reflections, the greatest intensity is at the point where the angle of incidence is equal to the angle of reflection – therefore, as is demonstrated in the graph, the brightest point occurs at the location (0, 8.5, 8). The diameter of the reflection is much smaller than in the diffuse model due to the coefficient s in the specular model equation. In the case of this particular plot, $s = 999$, resulting in a small spot on the surface. Although this a typical value for s , smaller values would result in a larger area with higher reflected intensities.

2.2.2 Phong shading model

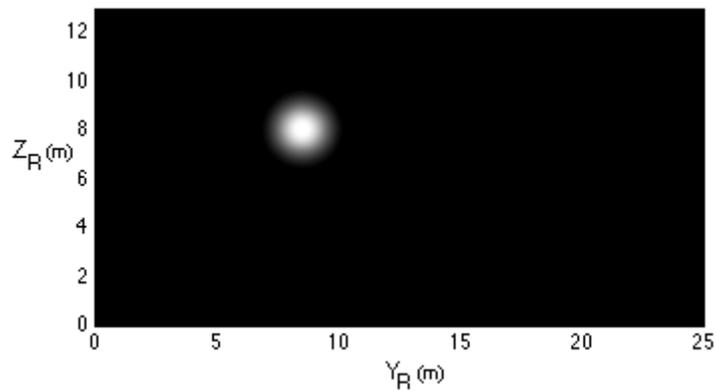


Figure 2.6 Specular reflection of a light source at coordinates specified in Figure 2.4 and a height of 8.

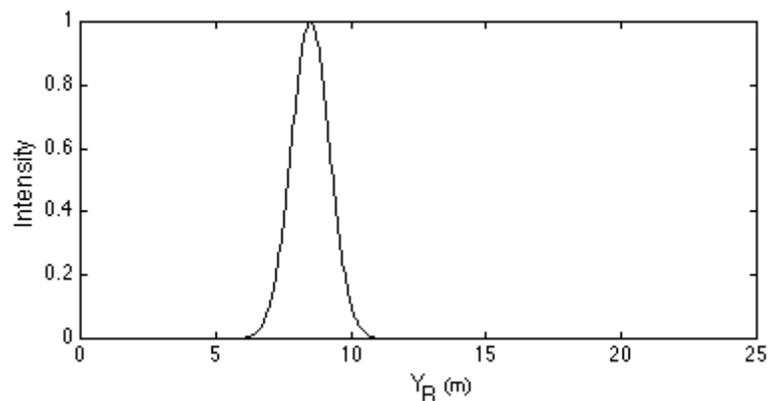


Figure 2.7 Plot of Intensity vs. location on specular reflecting wall model at $Z_R = 8$ m for locations shown in Figure 2.4 ($s = 999$).

If a slice of the plot in Figure 2.6 is taken along the line $Z_R = 8$ and the intensity of the reflectance is plotted for various values of Y_R , the graph in Figure 2.7 is the result. This shows not only that the highest intensity of the plot occurs at $Y_R = 8.5$, but also that the range of intensities is much greater than in the case of the diffusive model. Note that,

2.2.2 Phong shading model

whereas the diffuse model plotted in Figure 2.5 has a global minimum of approximately 0.78, the global minimum in the specular model in Figure 2.7 is 0. Again, this is a result of the high s coefficient.

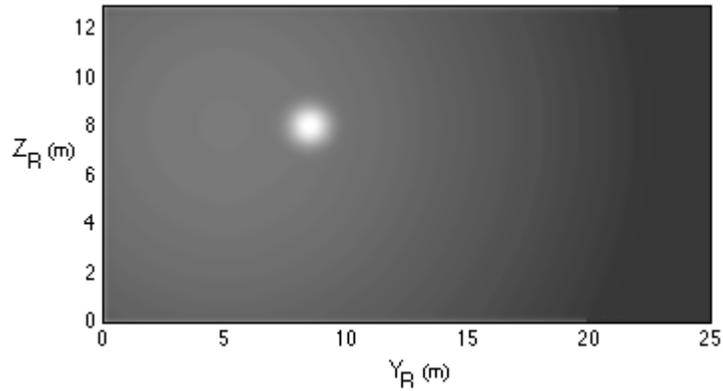


Figure 2.8 Diffuse and specular reflections of a light source at coordinates specified in Figure 2.4 and a height of 8.

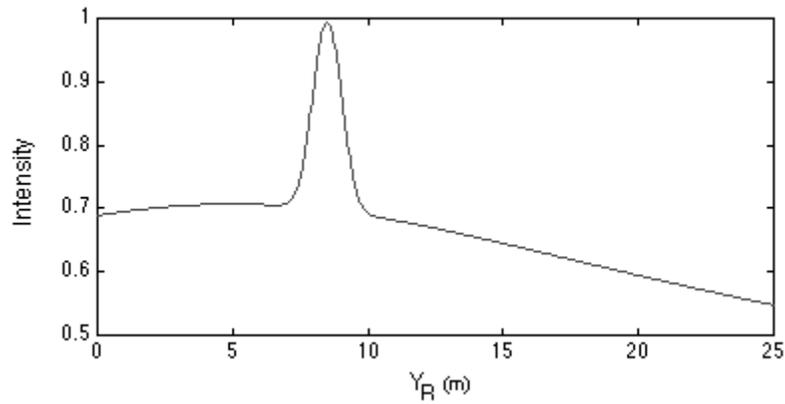


Figure 2.9 Plot of Intensity vs. location on wall model with 50% diffusive and 50% reflective properties for locations shown in Figure 2.4 ($s = 999$).

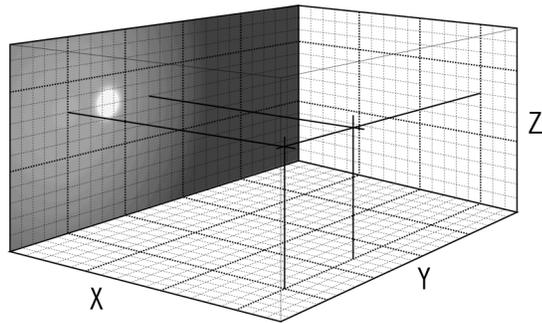


Figure 2.10 Three-dimensional representation of the plot in Figure 2.8 showing the locations of the light source at (15, 5, 8) and camera at (15, 12, 8).

3 Two-dimensional acoustic model

Whereas graphics models of virtual environments use methods to find the light intensity at a given point in space, virtual acoustic models are as concerned with the mapping of sound intensity over time as well as space. This issue of spatial distribution will be discussed in Chapter 4.

The model uses a combination of both empirical and physical models in order to simulate the response of a partially diffused first reflection. Two modified functions from the Phong shading algorithm are used, the Lambertian gain windowing used in diffuse reflections as well as the ability to have a user-controlled balance between the specular and diffuse components. Also, an empirical directivity function for sound sources is introduced. In addition, a physical model of the gain due to propagation, the local impedance response of a Schroeder diffuser, and the polar sensitivity patterns of zeroth- and first-order directional microphones are developed.

3.1 Phong shading model adaptation

As was illustrated in Section 1.7, typical modern reverberation engines use simple tapped delay lines to produce early reflection patterns. Although, as has been demonstrated, these delays may undergo a modicum of processing to simulate the absorptive or diffusive characteristics of a surface, this rarely, due to computational requirements and availability, extends beyond simple low pass filtering. This standard parallels the early graphics systems, the exception being that, whereas initial graphics assumed that all surfaces were diffusive reflectors, current reverberation engines assume that all surfaces are specular reflectors. To simulate a more realistic model, it is necessary to incorporate both the diffusive and specular components while offering some balance control of the two to the end-user.

In order to accomplish this, we will begin with a version of the Phong shading model described in Chapter 2. Some initial alteration is required for the model since an ambient energy source will not be included in the early reflection information. This ambient diffused energy is analogous to the reverberant decay which is generated using either a TRD model or large image model as described in Chapter 1. The procedure is therefore reduced to two components as follows.

Intensity of reflection = intensity of specular component + intensity of diffuse component

$$i_r = i_s + i_d \tag{3.1}$$

3.1.1 Specular component

Since the simple Phong shading algorithm was originally developed for a still image, it is able to rely on a number of basic assumptions. The most significant of these is that the light source is an infinite distance from the visual scene. This assumption both simplifies angle calculations and eliminates gain differences due to propagation. In order to further modify the model for acoustic sources in an enclosed environment with finite propagation distances, we must account for gain differences due to the total distance travelled by each reflection. The basic gain model used for this computation is shown in Equation 1.7 where the gain of the signal level is inversely proportional to the distance.

Because of the substantially different characteristics of the diffuse and specular components both of this model and of reality, the appropriate gain changes caused by propagation must be calculated differently for each. Specular reflections behave differently from their diffuse counterparts because, as is assumed in the traditional image model, the former is considered to be an exact copy of the original sound source, located on the opposite side of the reflecting surface. An illustration of this is an image seen in a mirror – it appears that a second light source exists behind the glass at a distance equal to that between the mirror and the real light source. We can therefore consider that, in the case of perfectly specular reflections, the pressure of the reflected copy of the sound source is dependent solely upon the total distance travelled by the sound wave. As a result, the function shown in Equation 3.2 can be used to determine the gain multiplier due to propagation of a specular reflection denoted G_{s2} .

3.1.1 Specular component

$$G_{s2} = \frac{1}{\overline{IR} + \overline{RM}} \quad (3.2)$$

where \overline{IR} and \overline{RM} are the distance from the instrument to the reflection point and the reflection point to the microphone respectively and can be calculated using Equations 3.3 and 3.4.

$$\overline{IR} = \sqrt{X_I^2 + (Y_I - Y_R)^2} \quad (3.3)$$

$$\overline{RM} = \sqrt{X_M^2 + (Y_M - Y_R)^2} \quad (3.4)$$

where the instrument is located at position (X_I, Y_I) , the microphone at location (X_M, Y_M) and the reflection point at $(0, Y_R)$.

This is an implementation of the basic image model with the formulae tailored to a reflective surface located on the Y-axis. If the surface is assumed to be a perfect reflector for all frequencies, then the filter is a delay line with the gain value G_{s2} applied to the output. The delay time D_s for a given reflection is also dependent on the propagation distance and can be calculated using Equation 3.5.

$$D_s = \frac{\overline{IR} + \overline{RM}}{c} \quad (3.5)$$

The transfer function of this filter is a simple attenuated and delayed impulse. However, it may be desirable to simulate the absorptive characteristics of the specular reflector as is implemented in commercial units. This can be accomplished by applying a filter with the desired characteristics to the output of the delay line, using either a simple filter, or a

3.1.1 Specular component

measurement of an actual reflection off a real surface. In either case, the delay inherent in the filter must be compensated in order to preserve the accuracy of the delay time calculation in Equation 3.5 above.

As in the graphics model, the level of the specular component of the reflection is controlled by a specular reflection scalar k_s which is multiplied by Equation 3.5 as shown in Equation 3.6.

$$G_s = k_s \frac{1}{\overline{IR} + \overline{RM}} \quad (3.6)$$

Note, however, in the graphics model, the sum of the specular and diffuse scalars was arranged to equal 1. This was implemented since the value calculated by the Phong shading model is based on the intensity, not the amplitude of the signal. Since the gain functions in the acoustic model are applied to the amplitude of the signal, the scalars must be modified in order to ensure that adjustments in the system result in an equal summed power. Consequently, rather than setting $k_s + k_d = 1$ as in the Phong shading model, the system requires that $k_s^2 + k_d^2 = 1$. The implementation used for all tests was based on the standard constant power panning curve and is shown in Equations 3.7 and 3.8.

$$k_s = \cos \left[\text{diffuse} \frac{\pi}{2} \right] \quad (3.7)$$

$$k_d = \sin \left[\text{diffuse} \frac{\pi}{2} \right] \quad (3.8)$$

3.1.1 Specular component

where *diffuse* is the power level of the diffused reflection component and $0 \leq \text{diffuse} \leq 1$.

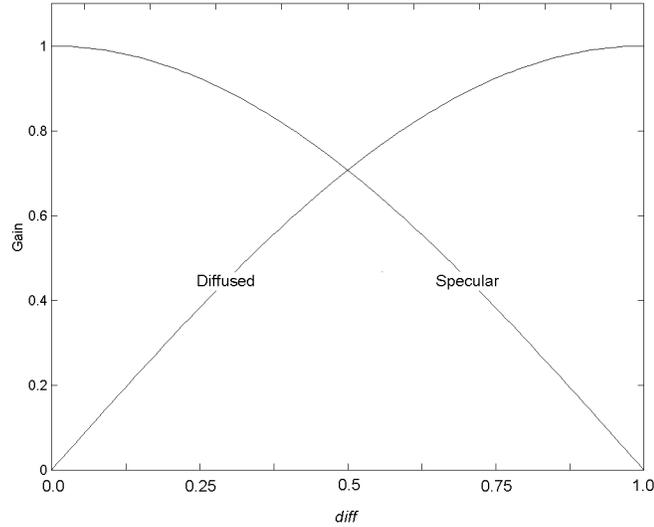


Figure 3.1 Relative gains of the diffused and specular components for various values of *diffuse*.

3.1.2 Diffused component

The gain of the diffused component G_{dl} due to the Phong shading model assumes a perfect Lambertian behaviour and is dependent both on the scalar k_d and on the cosine of the angle of incidence of the radiating sound wavefront as is shown in Equation 3.9.

$$G_{dl} = k_d \cos(\vartheta_i) \quad (3.9)$$

The angle of incidence ϑ_i required for this calculation can be computed given the locations both of the instrument (X_I, Y_I) and of the reflection point ($0, Y_R$) as shown in Figure 3.2 and expressed in Equation 3.10.

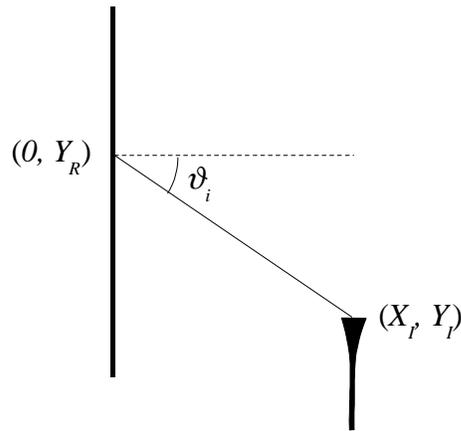


Figure 3.2 Relationship between instrument location (X_I, Y_I) , the point of reflection $(0, Y_R)$ and the angle of incidence ϑ_i .

$$\vartheta_i = \text{atan} \left[\frac{Y_R - Y_I}{X_I} \right] \quad (3.10)$$

Using this equation, and varying the location of the reflection point along the wall, it is evident that the intensity of the reflection undergoes a smooth change within a relatively limited gain range across the full length of the wall. Since this gain is dependent on the cosine of the angle ϑ_i , the maximum value for the gain occurs when the sound source is located on the normal to the surface through the reflection point, thereby resulting in an angle of incidence of 0 radians. The range of the angle, and therefore the gain, is dependent both on the length of the surface and on its distance from the sound source.

Figure 3.3 shows a Gain vs. Location plot for a perfectly diffusing surface 25 m in length situated along $X_R = 0$ with the sound source located 12 m from the wall at $(12, 5)$ as

3.1.2 Diffused component

shown in Figure 2.3. Note that the highest level, unity gain, is located at $Y_R = 5$, where the incident ray of sound is perpendicular to the reflecting surface.

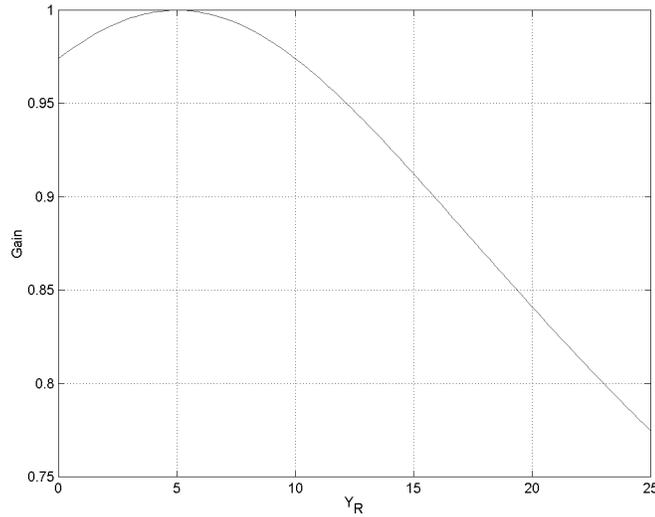


Figure 3.3 Gain of intensity relative to location on perfectly diffusing wall model for locations shown in Figure 2.3.

In the case of the diffused component, unlike that of the specular reflection, we are considering each reflection point along the wall's surface to be a new and independent sound source. Each of these points is a modified copy of the original sound source, with an intensity dependent upon the intensity of the instrument and its distance from the reflection point. As a result, the incident pressure at the point on the surface resulting

from attenuation solely due to propagation is $\frac{1}{IR}$. Since we are receiving the signal from this point at a second location – that of the microphone – the gain change caused by

3.1.2 Diffused component

the propagation from the reflection point to the microphone, $\frac{1}{RM}$ is multiplied to result in a propagation attenuation G_{d2} as shown in Equation 3.11.

$$G_{d2} = \frac{1}{IR RM} \quad (3.11)$$

This gain is, in turn, multiplied by the pressure gain converted from the shading model in Equation 3.9 to result in the final general gain function for the diffuse component G_d shown in Equation 3.12.

$$G_d = k_d \frac{\cos(\vartheta_i)}{IR RM} \quad (3.12)$$

The result of this function is a curve with a shape roughly similar to that of the simple Phong shading model with a number of significant differences. As can be seen in Figure 3.4, the global maximum for the curve has moved from the $Y_R = 5$ as was shown in Figure 3.3 to approximately $Y_R = 8.5$, similar to the specular reflection model. This is a result of the fact that the gain due to propagation is a more significant factor than the cosine multiplier when the source is distant from the surface, therefore the highest gain occurs near the point on the reflecting surface which corresponds to the shortest distance traveled by the sound wave regardless of its angle of incidence. The relative importance of these two factors will be further evaluated in Section 3.5.2.

3.1.2 Diffused component

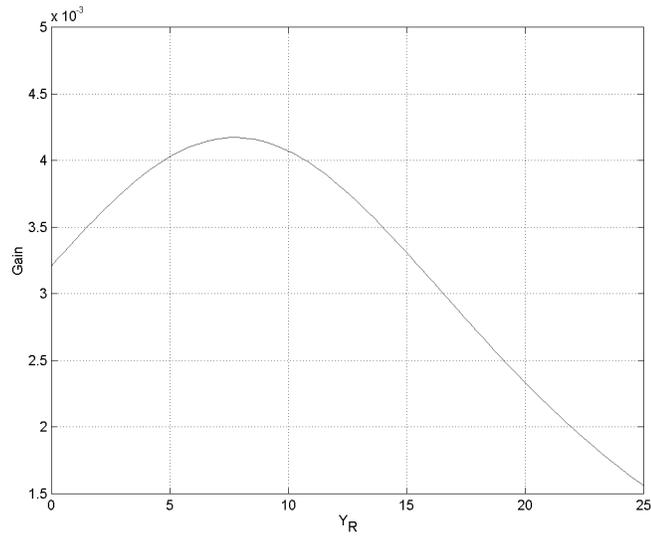


Figure 3.4 Gain of intensity relative to location including gain due to propagation on perfectly diffusive wall model for locations shown in Figure 2.3.

In addition, it is important to note the significant drop in gain due to propagation. This value is expected and will vary according to the particular spatial relationships of the source, microphone and reflecting surface.

3.2 Instrument directivity

It is common practice with external reverberation processors to feed the input with a so-called “dry” or anechoic signal that is mixed with the synthetic reverberant components at or after the output stage of the processor. With few exceptions, this signal is processed by the device with no attention paid to directional characteristics of the instrument. While the directivity of a musical instrument cannot easily be modeled using deterministic methods, we can allow for some simple control over the directional characteristics of the radiated power.

There are a number of auralization software packages that accommodate directivity patterns, typically using measurements of frequency- and distance-dependent polar patterns on regularly spaced angles. One such implementation has resulted in a suggestion for a standardized file format for directivity data known as the *DDI* or *DDL Directivity Index* (CATT 1998).

3.2.1 Instrument radiation patterns

In order to control the directivity pattern of the instrument, a function is used which provides a continuously variable gain which is dependent on the angle of the radiated sound wave. Since we are calculating the amplitude of the sound source at various discrete points on the reflecting surface, we can determine the change in level of the signal as a result of the angle from the instrument to the reflection point. This function must be variable such that the polar radiation pattern of the instrument can be modified by

3.2.1 Instrument radiation patterns

the user from a completely omnidirectional radiation through to a very narrow beaming effect. This can be accomplished using a function of the angle of radiation similar to one commonly seen in first-order microphone sensitivity polar patterns. This function, shown below gives the user a wide control over the directivity with a single variable g .

$$G_{\sigma_i} = [0.75 + 0.25 \cos(\sigma_i)]^g \quad (3.13)$$

where G_{σ_i} is the gain applied to the signal radiating in the direction σ_i , σ_i is the angle of radiation of the sound source, calculated using the function in Equation 3.14, and g is the directivity coefficient.

$$\sigma_i = \text{atan} \left[\frac{X_I}{Y_R - Y_I} \right] \quad (3.14)$$

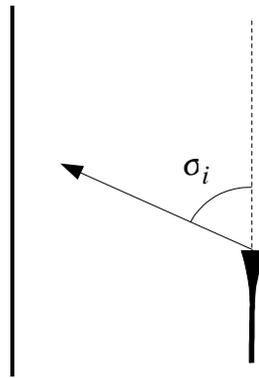


Figure 3.5 Angle of radiation of the instrument

3.2.1 Instrument radiation patterns

Changing the value of g can modify the radiation pattern of the sound source from completely omnidirectional (when $g = 0$) as in most commercial processors, through to a very narrow beam for large values of g (on the order of 20). To illustrate this effect, a number of radiation patterns are plotted below in Figure 3.6 for various values of g .

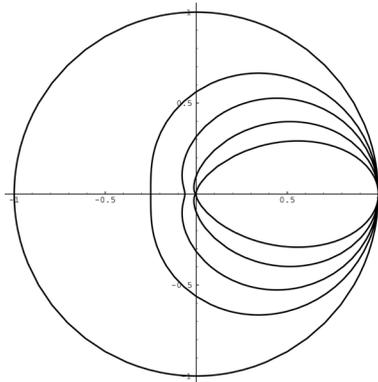


Figure 3.6 Sample polar plots of various instrument directivity curves for $g = \{0, 2, 4, 8, 16\}$ using Equation 3.13.

3.2.2 Instrument rotation

The inclusion of a process that simulates directional characteristics of the instrument subsequently requires that the user be able to rotate the instrument, pointing it in any direction. This rotation angle ζ_i is simply subtracted from σ_i in the above calculation as shown in Figure 3.7 and Equation 3.15.

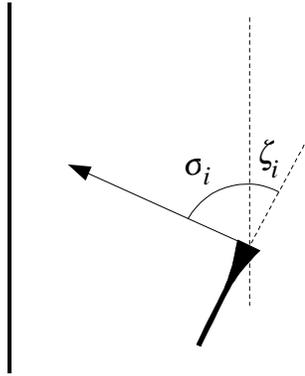


Figure 3.7 Angles of rotation and radiation of the instrument

$$G_{\zeta\sigma_i} = [0.75 + 0.25 \cos(\sigma_i - \zeta_i)]^g \quad (3.15)$$

where $G_{\zeta\sigma_i}$ is the gain applied to the signal directed toward the Reflection Point, σ_i is the angle subtended by the ray drawn between the instrument location and the reflection point and the line $x = X_I$, ζ_i is the angle of rotation of the instrument, and g is a directivity coefficient. (Note that positive changes in ζ_i indicate a clockwise rotation of the instrument when viewed from above.)

Figure 3.8 shows the effect of rotation of the instrument on the amplitude of the various reflections off points along the wall. As would be expected, the highest gain for the specular reflection occurs when the instrument is pointing towards the reflecting surface near the point of specular reflection at $\zeta_i = 270^\circ$.

3.2.2 Instrument rotation

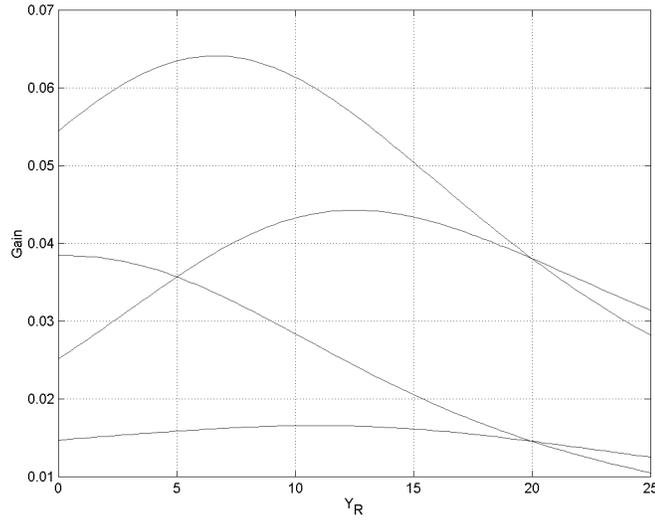


Figure 3.8 Instrument rotation (along the Y-axis: 270° top, 180°, 0°, 90° bottom). The plot includes gain changes due to the Phong model and propagation. $g = 2$.

Moore's general model for multichannel reverberation (Moore 1983) is one of the early documented synthetic acoustics engines to implement a directivity control for the sound source. In this paper, he suggests the function shown in Equation 3.16.

$$G_{\sigma_i} = \left[1 + \frac{(g-1)|\zeta_i - \sigma_i|}{\pi} \right]^2 \quad (3.16)$$

where G_{σ_i} is the scalar for radiation in the ray direction σ_i , g is a value of radiation from the rear of the instrument and is variable between 0 and 1, and ζ_i is the angle of rotation of the instrument.

While this function does provide control of the directivity pattern of the instrument, it suffers from a number of problems which are avoided in the proposed implementation.

3.2.2 Instrument rotation

Firstly, the directional pattern has a maximum directivity when the value of g is set to 0, resulting in a directivity pattern as shown in Figure 3.9. Furthermore, for all values of the g scalar other than 1, there is a discontinuity in the on-axis direction as can be seen in the sharp corners in the response's polar plot. Although this is not an issue for static instrument parameters as is the case with this particular implementation, it can cause audible artifacts in a dynamic environment.

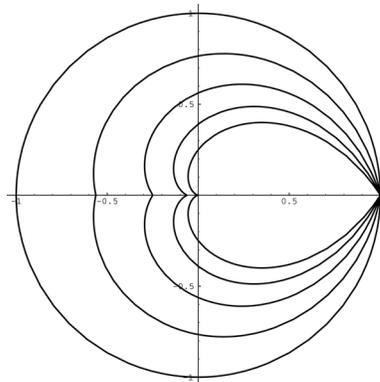


Figure 3.9 Sample polar plots of Moorer's instrument directivity functions for various values of $g = \{0, 0.25, 0.5, 0.75, 1\}$

It is important to note that both of these functions are intended to give an empirical representation of the directional characteristics of the instrument. One significant difference between the algorithms and radiation patterns typically seen in real instruments is the frequency-independence of the functions. Measurements of the radiation patterns both of real instruments (Meyer 1975) and of simple models of sound generators (Olson 1957) all show a tendency for the directivity to be proportional to the frequency emitted by the instrument. Very low frequencies tend to be almost omnidirectional and high

frequencies have a tendency to “beam,” depending on the relationship between the wavelength of the frequency in question and the size of the radiating body or horn. There will be two principal results of this simplification. The first is a lack of change in frequency response characteristics with respect to rotation in the instrument’s direct sound. This will affect both moving sources as well as different frequency response characteristics in spaced microphones. The second will be an error in the relative frequency response curves of the direct and reflected powers. Assuming that a directional instrument was positioned such that the microphone was on-axis, it is likely that the reflections off various surfaces in the room would be radiated off-axis to the instrument. As a result, there would be an expected loss of high-frequency information in the power directed towards the reflecting surfaces, thus increasing the direct to reflected level ratio at high frequencies.

As can be seen in Figure 3.10, the increased directivity of the instrument has the general effect of attenuating the specular and diffuse signals reflected off the various points on the wall (although this behavior is dependent on the angle of rotation of the instrument). In addition, it changes the relative balance of the amplitude of the reflections off the various points on the wall. Note that, in the case of the most directional instrument setting plotted ($g = 16$), the diffuse reflection off the wall at $Y_R = 25$ has a higher gain than the specular component at $Y_R = 8.5$. When the instrument is omnidirectional at $d = 0$, the opposite is true. It should also be pointed out that the higher gain “compression” seen at $Y_R = 25$ exists due to the fact that the instrument is aimed towards $Y_R = 5$, therefore the

3.2.2 Instrument rotation

change to a higher directivity will have less of an effect towards the front of the instrument

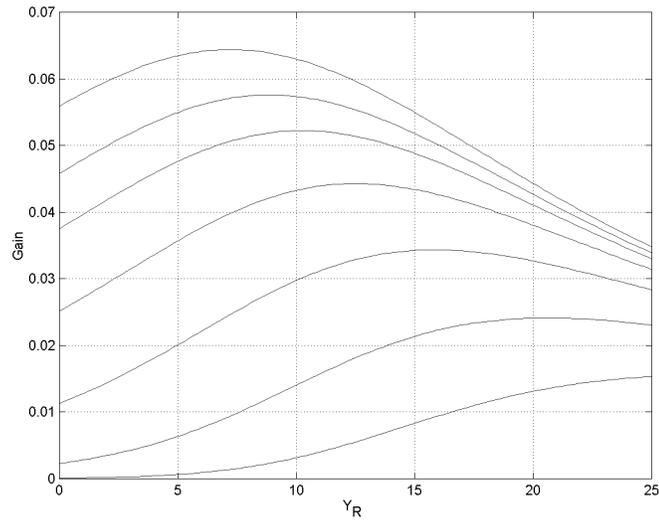


Figure 3.10 Gain responses for various values of g for the instrument directivity. $g = \{0$ (top), $0.5, 1, 2, 4, 8, 16$ (bottom)}. $\zeta_i = 0^\circ$. The plot includes gain changes due to the Phong model and propagation.

3.3 Microphone sensitivity

As was discussed in the introduction, the present incarnation of the SceneBuilder software includes the option for the user to locate virtual microphones in the synthetic acoustic environment. These microphones can not only be positioned anywhere within the boundaries of the enclosure, but have adjustable directional characteristics and rotation.

3.3.1 Microphone polar patterns

The directional characteristics of microphones are both more familiar and easier to implement than instrument directivity. Much has been written regarding the sensitivity patterns of first-order microphones using the standard sensitivity function shown in Equation 3.17 (Woram 1989).

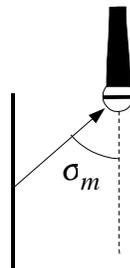


Figure 3.11 Diagram showing the angle of incidence σ_m for the microphone

$$G_{\sigma_m} = P_m + (1 - P_m) \cos(\sigma_m) \quad (3.17)$$

where G_{σ_m} is the sensitivity of the microphone at angle of incidence σ_m and P_m is the pressure (or omnidirectional) component of the transducer. Note: Typically, the pressure

3.3.1 Microphone polar patterns

gradient component is labeled “ PG ” or equivalent. Its value, however, is always equal to $(1 - P_m)$, hence the notation used.

Again, in order to calculate the sensitivity of the microphone at a given angle of incidence the angle of incidence must be calculated using Equation 3.18.

$$\sigma_m = \text{atan} \left[\frac{X_M}{Y_M - Y_R} \right] \quad (3.18)$$

Figure 3.12 shows the plots for three different sensitivity patterns for the microphone in a stationary location.

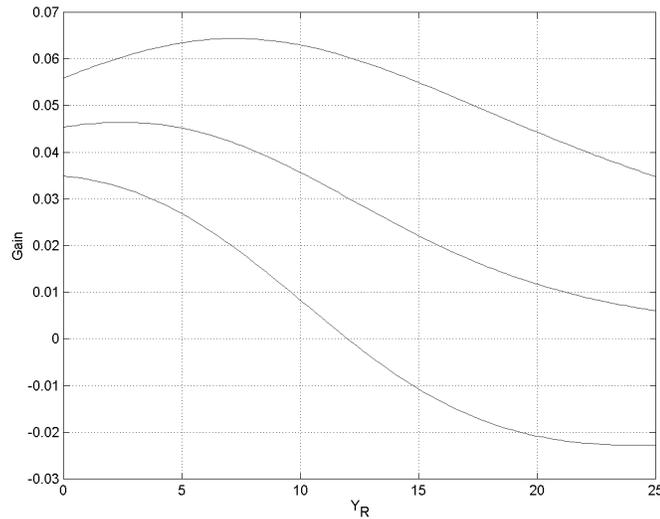


Figure 3.12 Reflection gains for various microphone polar patterns including Phong model and gain due to propagation. Top – omnidirectional ($P_m = 1$), Middle – cardioid ($P = 0.5$), Bottom – bidirectional ($P = 0$). Note the negative gain for the rear lobe of the bidirectional microphone ($g = 0$).

3.3.2 Microphone rotation

As in the case of the instrument directivity, the inclusion of a directional characteristic for the microphone necessitates a capacity to control the angle of rotation of the microphone.

This angle ζ_m is added to the angle σ_m resulting in Equation 3.19.

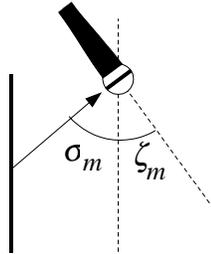


Figure 3.13 Diagram showing angles of incidence and rotation for the microphone

$$G_{\zeta\sigma_m} = P_m + (1 - P_m) \cos(\zeta_m + \sigma_m) \quad (3.19)$$

where $G_{\zeta\sigma_m}$ is the gain applied to the signal arriving at the microphone from the direction of the Reflection Point, σ_m is the angle subtended by the ray drawn between the microphone location and the reflection point and the line $x = M_x$, and ζ_m is the angle of rotation of the microphone. Note that positive changes in ζ_m indicate a clockwise rotation of the instrument when viewed from above.

3.3.2 Microphone rotation

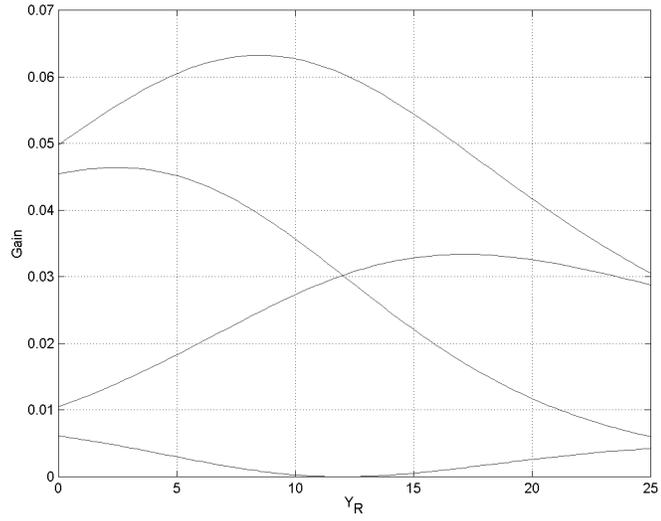


Figure 3.14 Gain plot for reflections at values of Y_R for 4 rotations of a cardioid microphone ($P_m = 0.5$). Along the Y-axis: 90° Top, 0°, 180°, 270° Bottom. ($g = 0$).

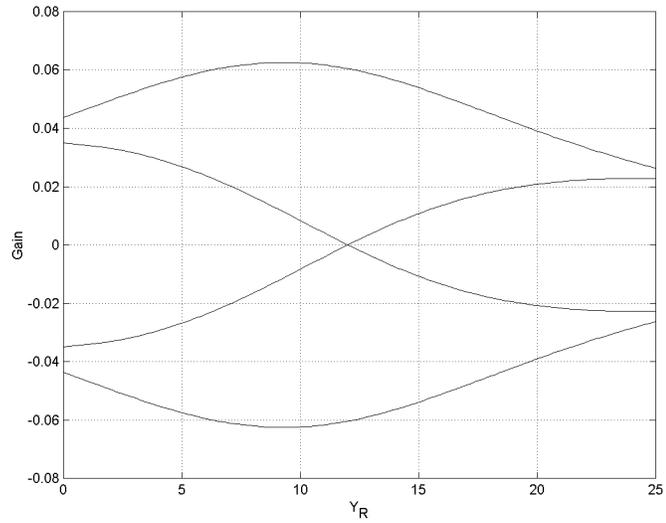


Figure 3.15 Gain plot for reflections at values of Y_R for 4 rotations of a bidirectional microphone ($P_m = 0$). Along the Y-axis: 90° Top, 0°, 180°, 270° Bottom. ($g = 0$).

3.4 Modelling wall details

The gain of the diffuse component for various locations along the wall's surface using the formulae described in Sections 3.1 and 3.2 provide only a macroscopic envelope for both the spatial and temporal characteristics of the impulse response. While this envelope provides the end user with parameters that determine the relative contributions of energy from various locations along the reflecting surface, they are used simply as a location- and therefore time-dependent gain which is applied to the local impulse responses along the surface of the diffuser. In the case of a Schroeder diffuser, this local impulse response matches the impulse response of a reflection from the mouth of a single well.

3.4.1 Schroeder diffuser wells

As has been previously stated in Section 1.2.2.3, a Schroeder diffuser can be considered as a flat reflective surface with an impedance which varies according to location. This local impedance is dependent upon the width and depth of each individual well in the diffuser and can be calculated using Equation 1.22. Although the plot in Figure 1.12 shows a graph of the reactance vs. frequency for a well of specific dimensions, it does demonstrate the characteristic shape of the impedance response for any well with a bottom with an infinite acoustical impedance. Note, however, that in real cases, since the well bottom has a finite acoustic impedance, the resulting impedance of the well mouth is comprised of both an acoustic resistance and reactance. An example showing both of these frequency-dependent components is plotted in Figures 3.16 and 3.17. The sample used for these calculations, using Equation 1.22 assumes a diffuser well with a depth of

3.4.1 Schroeder diffuser wells

8.6 cm, a width of 4.71 cm and a cap with an acoustic impedance matching solid oak at 1 kHz. Note that the lowest value in the acoustic resistance plot in Figure 3.16 is equal to the resistance of the construction material for the well bottom.

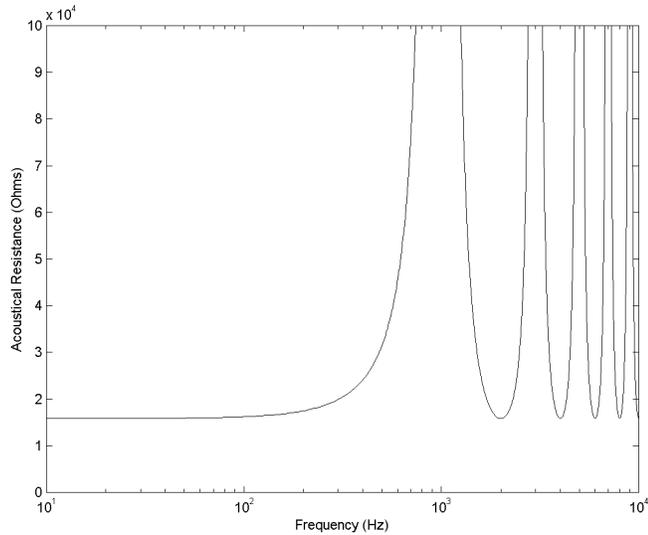


Figure 3.16 Calculated real component of the impedance vs. frequency for a diffuser well of depth 8.6 cm and width 4.71 cm (design frequency = 1000 Hz) and a circular cross section. Acoustic impedance of well cap equivalent to measured value of oak at 1 kHz.

3.4.1 Schroeder diffuser wells

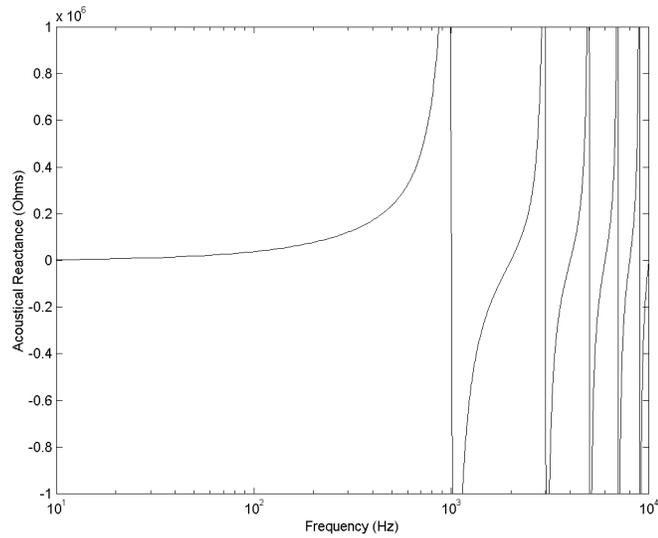


Figure 3.17 Calculated imaginary component of the impedance vs. frequency for a diffuser well of depth 8.6 cm and width 4.71 cm (design frequency = 1000 Hz) and a circular cross section. Acoustic impedance of well cap equivalent to measured value of oak at 1 kHz.

It should be noted that the impedance of the well bottom is predominantly determined by the absorptive characteristics of the construction material. If we assume that the absorption coefficient α is frequency-independent, then the pressure reflection coefficient R can be calculated using Equation 3.20 (Kinsler 1982).

$$R = \sqrt{1 - \alpha} \quad (3.20)$$

The relationship between the impedance of the reflecting surface and the pressure reflective coefficient is shown in Equation 1.13.

Since, in the case of a progressive plane wave, the acoustic impedance of a material contains only a real component, then r_l in Equation 1.13 is equal to the characteristic impedance of air, a value of 416.24 ohms. Consequently, if we assume a frequency-independent absorption, the impedance of a material can be calculated using Equation 3.21.

$$z_d = 416.24 \left[\frac{1 + \sqrt{1 - \alpha}}{1 - \sqrt{1 - \alpha}} \right] \quad (3.21)$$

In order to determine a predicted impulse response for an individual well, this impedance function must first be converted from the frequency to the time domain using an Inverse Fast Fourier Transform of the impedance response as in Equation 3.22.

$$IFFT \left[\frac{\frac{\rho_o c}{S_n} \frac{z_d + j \frac{\rho_o c}{S_n} \tan(kd)}{\frac{\rho_o c}{S_n} + j z_d \tan(kd)} \right] \quad (3.22)$$

This IFFT converts results to a time domain representation of the impedance response of the mouth of a diffuser well as measured from the room. Figure 3.18 shows a plot of such a representation for the impedance vs. frequency graphs in Figures 3.16 and 3.17.

3.4.1 Schroeder diffuser wells

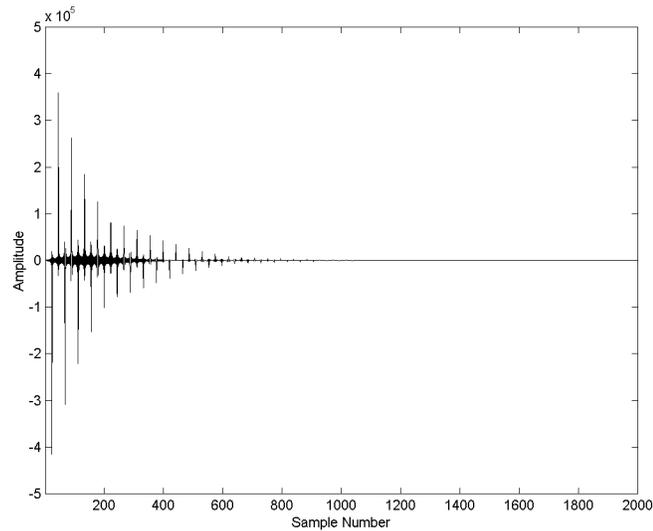


Figure 3.18 Time domain representation of the combined real and imaginary components of the impedance response plotted in Figures 3.16 and 3.17.

Although the result of the IFFT is a time domain representation of the impedance response of the well mouth, it must be further modified in order to be used in convolution with the proposed system. If a recording of an audio signal's pressure wave were convolved through this impedance response the resulting output would be a simulation of the reflected velocity wave. As was explained in Section 1.1, this is because an acoustic impedance is the product of the acoustic pressure and the reciprocal of the particle velocity. Consequently, the output of the system must be converted back to a representation of a pressure wave before the system is complete. Since a velocity component of an acoustic wave is the first derivative of its pressure component, this can be accomplished by convolving the velocity signal with a first-order difference equation, approximating a derivation filter. If we consider the velocity wave to be the function $f(x)$,

the first derivative of the function can be approximated using Equation 3.23 (Van Duyne 1993).

$$f'(x) = \frac{f(x + \Delta x) - f(x)}{\Delta x} \quad (3.23)$$

This approximation can be expressed as a digital filter using a single unit delay as in Equation 3.24.

$$y[n] = \frac{x[n+1] - x[n]}{T_s} \quad (3.24)$$

where T_s is the sampling period of the system.

Rather than convolve the audio signal through the derivation filter, it is more efficient to compute instead the derivative of the time domain representation of the impedance response of the well before the signal is convolved through it. Figure 3.19 shows the result of the impulse response in Figure 3.18 above filtered with the first-order difference equation.

3.4.1 Schroeder diffuser wells

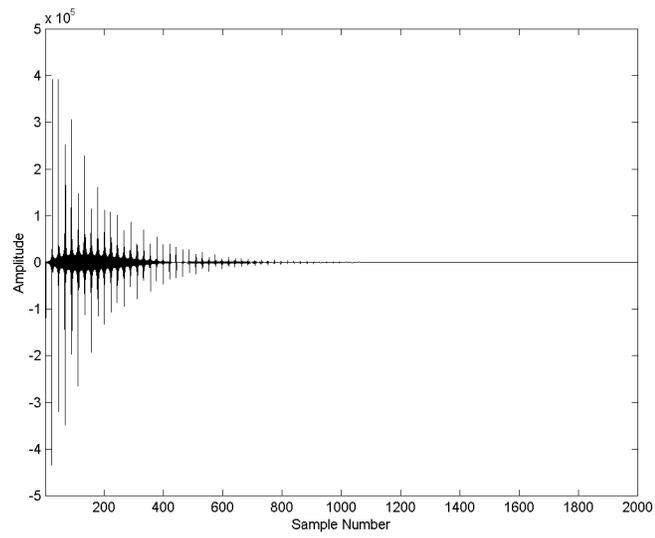


Figure 3.19 Impulse response of entrance of well mouth from room, produced by calculating the first derivative of the time domain representation of the impedance response plotted in Figure 3.18.

3.5 Building the full impulse response

The impulse response of a local reflection off a given surface is calculated using the four components discussed in Sections 3.1 through 3.4. A wall of length L is assumed to be located on the Y-axis of a Cartesian coordinate system stretching from $(0, 0)$ to $(0, L)$. The microphone is located at (X_M, Y_M) and the instrument at (X_I, Y_I) . Given these parameters, the point of reflection will be located at a point (X_R, Y_R) on the surface where $X_R = 0$ and $0 \leq Y_R \leq L$.

3.5.1 Specular reflection component

The specular component is comprised of a gain level written at a single sample in the impulse response. The time index of this sample is calculated using the image model as described in Section 3.1.1 and Equation 3.5 and rounded to the nearest sample. The value of the amplitude is found by multiplying the results of Equations 3.6, 3.19 and 3.15 as shown in Equation 3.25.

$$G_{specular} = \left[\frac{k_s}{\overline{IR} + \overline{RM}} \right] \left[P_m + (1 - P_m) \cos(\zeta_m + \sigma_m) \right] \left[0.75 + 0.25 \cos(\sigma_i - \zeta_i) \right]^g \quad (3.25)$$

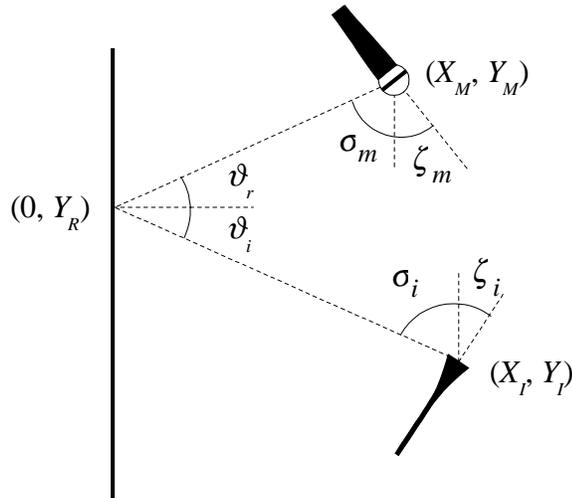


Figure 3.20 Relationship between the instrument and microphone locations as well as the point of specular reflection.

Although the specular component of the total reflection impulse response is comprised of only one sample, it is significant since it is the first non-zero sample in the reflection's impulse response. This is due to the fact that the specular reflection occurs from the point on the wall where the total propagation of the reflection is smaller than from any other point on the surface.

3.5.2 Diffused reflection component

The diffuse component of the reflection is calculated on a sample-by-sample basis, beginning with the earliest sample in the impulse response of the early reflection. The point of reflection on the surface corresponding to this component of the impulse response is the same point at which the perfect specular reflection occurs. As a result, its location and therefore time is most easily calculated using the image model. The local

impedance at this location is found by determining the dimensions of the diffuser well located at that point. Subsequently, the first derivative of the time response of this impedance is found using the procedure described in Section 3.4.1. This calculated impulse response for the local well (similar to that plotted in Figure 3.19) is globally multiplied by the gain modifiers derived from the Phong algorithm and directivities of microphone and instrument. This modified response is then written into the total impulse response for the entire reflecting surface.

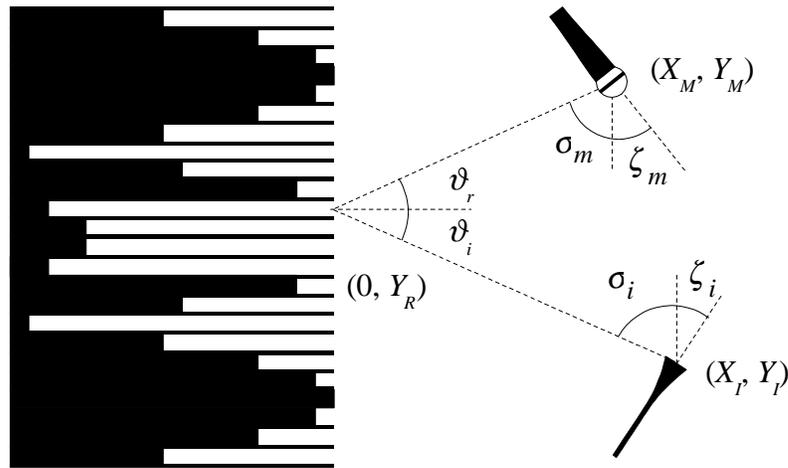


Figure 3.21 First diffuse reflection component

It must be considered that the single diffuse reflection results in multiple simultaneous reflections being received by the microphone. In the case of a one-dimensional reflecting surface, that is to say, one without height as is described in this model, this simultaneity is reduced to two reflection locations on either side of the point of specular reflection for a given instrument and microphone location. Unfortunately, a typical ray-trace methodology would not produce acceptable results in this case, since a discretized space,

either as regularly spaced angles around the instrument or microphone or reflection locations along the surface will produce aliasing artifacts due to conflicts with the discrete time of the calculated digital impulse response. This is a similar situation to that encountered in graphics modelling software where images at the camera location must be calculated on the basis of regularly-spaced pixels rather than regularly-spaced angles.

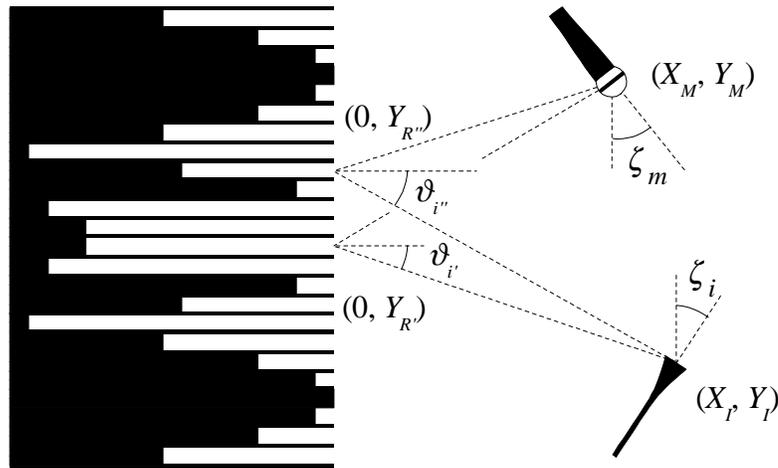


Figure 3.22 Simultaneously arriving reflections from two distinct locations on the surface. Note that these two points of reflection, denoted $(0, Y_R)$ and $(0, Y_{R''})$ have distinct angles of incidence, and therefore gain multipliers from the Phong algorithm. In addition, although not denoted on the diagram, they also result in distinct angles of radiation and incidence for the instrument and microphone respectively.

As a result, the location of the reflections after the initial onset of the impulse response must be calculated based on their time of arrival at the microphone. Based on the established convention of a microphone location at (X_M, Y_M) and an instrument located at (X_I, Y_I) with a reflecting surface on the Y-axis, the total distance of propagation D_{tot} of a

3.5.2 Diffused reflection component

given reflection can be derived from its time of arrival t (measured in samples) using Equation 3.26.

$$D_{tot} = \frac{c t}{F_s} \quad (3.26)$$

The total propagation distance for any reflection can be calculated using Equation 3.27

$$D_{tot} = \sqrt{X_I^2 + (Y_R - Y_I)^2} + \sqrt{X_M^2 + (Y_M - Y_R)^2} \quad (3.27)$$

This equation can be solved for Y_R resulting in Equation 3.28

$$\begin{aligned} Y_R = & (4D_{tot}^2 Y_I - 4X_I^2 Y_I - 4Y_I^3 + 4Y_I X_M^2 + 4D_{tot}^2 Y_M + 4X_I^2 Y_M + 4Y_I^2 Y_M - 4X_M^2 Y_M + 4Y_I Y_M^2 - \\ & 4Y_M^3 \pm \text{sqrt}[(-4D_{tot}^2 Y_I + 4X_I^2 Y_I + 4Y_I^2 - 4Y_I X_M^2 - 4D_{tot}^2 Y_M - 4X_I^2 Y_M - 4Y_I^2 Y_M + 4X_M^2 Y_M - \\ & 4Y_I Y_M^2 + 4Y_M^2)^2 - 4(4D_{tot}^2 - 4Y_I^2 + 8Y_I Y_M - 4Y_M^2)(-D_{tot}^4 + 2D_{tot}^2 X_I^2 - X_I^4 + 2D_{tot}^2 Y_I^2 - 2X_I^2 Y_I^2 \\ & - Y_I^4 + 2D_{tot}^2 X_M^2 + 2X_I^2 X_M^2 + 2Y_I^2 X_M^2 - X_M^4 + 2D_{tot}^2 Y_M^2 + 2X_I^2 Y_M^2 + 2Y_I^2 Y_M^2 - 2X_M^2 Y_M^2 - \\ & Y_M^4)]) / (2(4D_{tot}^2 - 4Y_I^2 + 8Y_I Y_M - 4Y_M^2)) \end{aligned} \quad (3.28)$$

Note that Equation 3.28 produces two results. This is due to the fact that two reflection locations, one on either side of the specular reflection location, will produce the same propagation distance and therefore time of arrival. This equation is used for each individual sample following the initial onset in the impulse response to find the local impedance response from which is derived a local impulse response which is subsequently added to the total response of the surface. This procedure is repeated until the end of the reflecting surface is reached, on one end at $(0, 0)$ and on the other at $(0, L)$.

3.5.2 Diffused reflection component

If we consider only the results of the modified Phong shading algorithm without the additional time and gain response of the results of the local impulse response of individual wells, a general gain model can be shown which is based on the interrelated time and spatial domains. Figure 3.23 shows the gain vs. sample number for a perfectly diffuse surface including the gain due to propagation for instrument and microphone locations matching the light and camera locations shown in Figure 2.3.

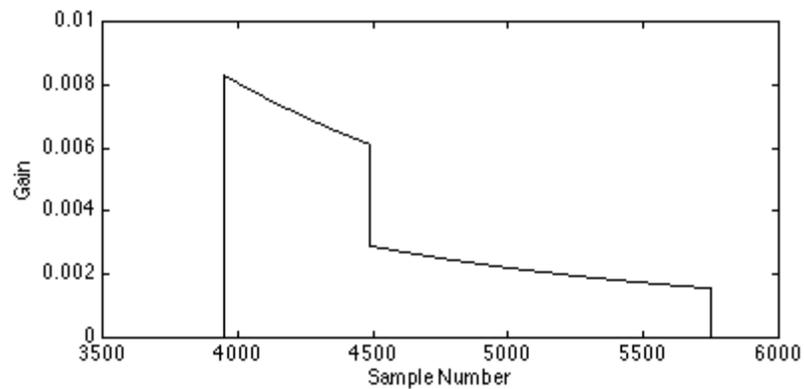


Figure 3.23 General gain response of a perfectly diffuse surface including shading algorithm and gain due to propagation ($F_s = 44.1$ kHz).

There are a number of interesting characteristics in this envelope which deserve discussion. Firstly, the relationship between the initial onset and the subsequent sample in the impulse response should be noted. Since the “attack” of the reflection is comprised of a single location on the reflecting surface whereas the subsequent sample is the result of the sum of the two adjacent reflections, there is effectively a 2-sample long attack time

for the curve. This response is shown in Figure 3.24 demonstrating that, in this particular arrangement, sample number 3950 is almost twice the value of sample 3949.

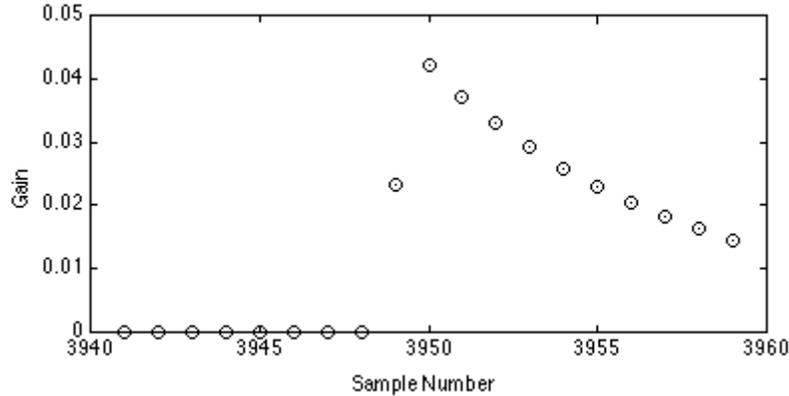


Figure 3.24 Detail of onset of gain envelope of perfectly diffuse reflection ($F_s = 44.1$ kHz).

A more obvious characteristic of the gain function is the sudden drop in level approximately half-way through the impulse response. This is due to the transition between the gain resulting from the sum of two reflections to an area where only one reflection arrives at the microphone. As was previously pointed out, this is the time at which one of the directions of propagation of the reflection point along the wall reaches the end of the reflecting surface.

Finally, it should be noted that, although the specific exponential decay of the gain curve shown in Figure 3.23 is a shape which can be assumed in most cases, it is possible to have an arrangement which results in a slightly different effect. The decay of the plot in Figure 3.23 is the product both of the modified Phong algorithm and of the propagation distance of the reflection. When the instrument and the microphone are relatively distant

from the reflecting surface, the gain due to propagation plays a significant role in this effect. If, however, the instrument is located near the reflecting surface, the Phong algorithm has an increased influence on the general shape, resulting in a smoother attack in the impulse response as is demonstrated in Figure 3.25. In this case, the instrument is located at (1, 5) and the microphone at (1, 12) with a reflecting surface located on the Y-axis from (0, 0) to (0, 25).

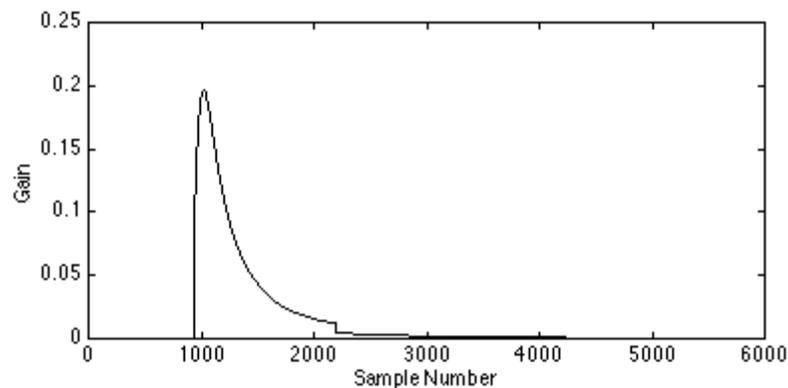


Figure 3.25 Gain function of a perfectly diffuse reflector including gain due to propagation with the instrument at (1, 5) and the microphone at (1, 12). The reflecting surface is located on the Y-axis and is 25 m long. Locations are set closer to the wall to emphasize the effect of the Lambertian behaviour on the response ($F_s = 44.1$ kHz).

3.5.3 Combining the specular and diffuse components

The system can be used in two implementations, depending on the speed of the processing system. Where real-time implementation is possible, then the relative balance

between the k_d and k_s scalars can be used to determine the balance between the specular and diffuse components in a single impulse response. This combined response would then be convolved with the anechoic sound file resulting in a reflection with the desired relative levels between the specular and diffuse characteristics.

Unfortunately, the current implementation does not permit immediate results with changes in the relationship between the two scalars. As will be discussed in Chapter 4, the prototype system used for testing the algorithm is implemented in MATLAB and takes a period on the order of hours to produce short listening samples on the order of seconds. Consequently, the system is used to produce three related parallel sound stimuli. The first is a direct sound, with each microphone's channel containing a single impulse with a delay time and gain determined by the direct distance between the instrument and the microphone. The second signal is based on a perfectly specular reflection model, with a single sample impulse for each reflecting surface in each microphone's output channel. The third and final sound file uses a perfectly diffuse model with user-defined diffuser design frequencies. These three sound files are then played in parallel with appropriate mixing of the various components done in real time using gain changes in the playback software.

It should be pointed out that this method serves simply to avoid the computational limitations of the present system. A mixing system based on the k_d and k_s scalars is preferable, but regrettably presently unfeasible except in emulation.

3.6 Advantages and disadvantages of the system

The original intention of the proposed system was to develop a method of diffusing early reflections that was based on the physical behaviour of real acoustic environments, but with the parameter controls of a perceptual model. While the system is successful in this regard, it is subject to a number of disadvantages and limitations.

Despite the fact that the system uses a physical model to produce the local impulse response characteristics of each individual diffuser well, the overall model cannot be considered to produce a result indicative of an actual space constructed of Schroeder diffusers. As a result, the system is inappropriate for predictive acoustics models and should not be used to analyze the reflection characteristics of real surfaces.

The principal advantage of the system is that it provides the end user with a manipulatable balance between perfectly specular and diffuse reflection models. Listener preferences within this balance was tested using psychoacoustic listening test procedures and will be evaluated in Chapter 5. This balance control permits end users to tailor the response of the system to different program material. In addition, the system can be used to model acoustic situations which would not be possible in real spaces. One example of this is the construction of Schroeder diffusers with extremely low design frequencies which would, in the real world, require prohibitive areas and volumes.

4 Implementation and analysis

Twenty years ago, the development of various digital reverberation engines was accomplished using fast convolution occupying hours of computer time on substantial mainframes. For example, in his 1979 paper, Moorer states that a time of 6 minutes was required to process each second of music through Schroeder's algorithms with 16 bit resolution using fast convolution implemented on a DEC PDP-10 computer. As processing speed and bandwidth increased over the years that followed, it became feasible to implement these algorithms as real-time processes on individual dedicated DSP integrated circuits. Today's "average" studio processors based on TRD models use on the order of 40-60 allpass and recursive delays in addition to multitaps with independent gains, all running in real-time at reasonably low expense.

Due to the non-recursive nature of the procedures described in Chapter 3 (with the exception of the well model) and in order to simplify implementation, the described system is implemented using convolution. The impulse response resulting from each instrument / reflective surface / microphone relationship is created individually using custom MATLAB functions and subsequently added, resulting in a total of four first-order reflections for each microphone collected in a single MATLAB matrix. This impulse response is then convolved with an anechoic, 16-bit, 44.1 kHz recording, producing a vector which corresponds to a single audio channel. If a gain reduction is necessary to avoid distortion, the matrix is normalized to a maximum amplitude of 1 with appropriate matching gain compensation applied to all other channels. This procedure is

repeated for various other microphone locations to generate multiple matrices, each corresponding to an intended output channel. These can then be used to create a single multichannel matrix which is subsequently exported as a 16-bit linear, NeXT / Sun format “.au” file. For reasons of compatibility between MATLAB and the playback software, this format was used for import and export of audio data. This file format is capable of a maximum of 16 bits of resolution with multiple channel compatibility.

4.1 Gain error analysis

The gain resolution of the audio files and impulse responses is subject to the accuracy of MATLAB's computations. These are done in double precision, resulting in a resolution equivalent to 64 bits (MathWorks 1998). This high resolution of MATLAB ensures a theoretical dynamic range of approximately 380 dB for each of the samples calculated in the impulse response. Note as well that the sampling rate of the files is independent of MATLAB and restricted only by the limitations of the file format used to import and export the data.

4.2 Propagation error and phase response

The procedure described in Chapter 3 for generating the impulse response of the reflection off a given surface incurs errors in the delay time calculations. In the description, quantization, rather than interpolation, is used to determine the appropriate sample in the impulse response to which a particular gain value is assigned. The magnitude of this incurred error is dependent on the delay resolution dictated by the

sampling rate chosen by the end user, and has a range of $\pm \frac{T_s}{2}$ where T_s is the sampling period. The error caused by this rounding effectively alters the distance to the reflective surface from a calculated value to a point closer to or further from the instrument and microphone. Since the delay is a consequence of the two-way trip to and from the reflection point, the error in terms of distance to the reflective surface is equivalent to one half the propagation distance of a sound wave in the time value of the error, or

$$E_d = \pm \frac{c}{4F_s} \quad (4.1)$$

where E_d is the delay error measured as a propagation distance in m, c is the speed of sound in m/s and F_s is the sampling rate in Hz.

Table 4.1 shows a list of maximum incurred errors for various sampling rates. As can be seen, these values are extremely small in relation to our ability to discriminate both

4.2 Propagation error and phase response

distance to the source or reflective surface and the typical distances from the surface to the instrument and microphone.

| Fs (kHz) | Maximum E_d |
|----------|----------------|
| 44.1 | ± 1.950 mm |
| 88.2 | ± 0.975 mm |
| 48 | ± 1.792 mm |
| 96 | ± 0.896 mm |

Table 4.1 List of propagation errors due to temporal quantization at various sampling rates

Although these values appear small, they may not necessarily be dismissed. The error caused by rounding the delay time to the nearest sample incurs a change in phase of the signal of differing amounts according to the frequency. Figure 4.1 shows a plot of the maximum phase error which can result from this quantization. As can be seen, the maximum error, occurring at the Nyquist frequency, is 90° .

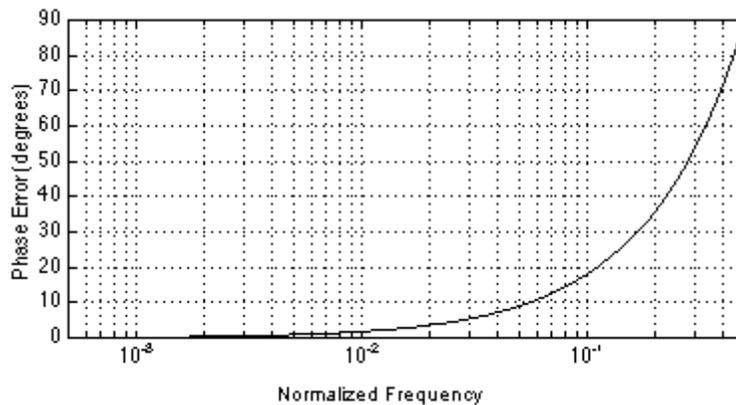


Figure 4.1 Maximum phase error vs. Frequency due to time delay quantization assuming that no interpolation or oversampling is employed

In order to reduce the effects of this quantization error, interpolated delays are required in the creation of the impulse response. This is achieved in MATLAB by oversampling the impulse response and resampling down to the sampling rate of the audio file before convolution. The procedures for computing the impulse response are identical to those described in Chapter 3, however, these are done using a multiple of the sampling frequency determined by the end user. As a result, the quantization error is reduced by a factor of the oversampling multiplier. Using the RESAMPLE function in MATLAB, the various computed delays in the impulse response are interpolated to the new sampling rate using a lowpass FIR digital filter and a Kaiser windowing function with compensated filter delay (MathWorks 1998).

4.3 Creation of sample sound files for measurement and listening tests

In order to test the system, using both electroacoustic measurements and psychoacoustic listening tests, sample sound files were required. These were created by convolving anechoic recordings through impulse responses which had been created using the described system. For the purposes of this dissertation, three monophonic sound files from the Bang & Olufsen test disc *Music for Archimedes* were used (Bang & Olufsen 1992), each chosen to provide a unique characteristic. These three recordings are of solo xylophone, solo cello, and female speech. The first of these was chosen to highlight the transient behaviour and high frequency response characteristics of the system, the second to test the steady state and low frequency response characteristics, and the third to highlight any possible differences between the simultaneous transient and steady state characteristics of the reflection model. The last was chosen also because it is a non-musical source. Appendix B contains specific information regarding the technical details of the sound recordings.

The room dimensions of the model used in the creation of the impulse response match those of McGill University's Redpath Hall, measuring 27.23 m long and 12.90 m wide. The virtual instrument was placed at a typical location for performers in Redpath when the stage is set against the West wall, 4.19 m from the West wall and 5.45 m from the South wall. The centre of the virtual microphone array was set at a typical location as well, 8.75 m from the West wall and 5.45 m from the South wall. Figure 4.2 shows a diagram of this configuration.

4.3 Creation of sample sound files for measurement and listening tests

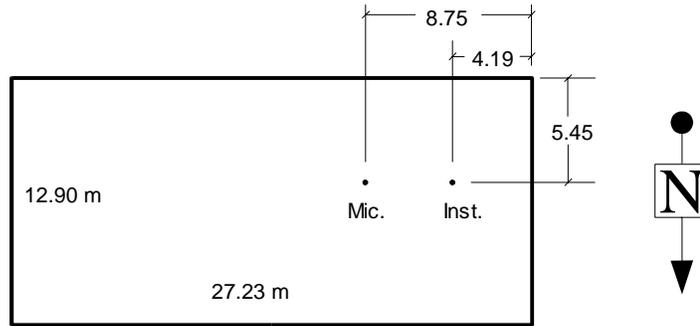


Figure 4.2 Room dimensions, microphone array and instrument locations used for the impulse responses created for audio tests.

The instrument was set to a directivity of 0, or completely omnidirectional. The microphones were arranged in a seven-channel array based on a “Fukada Tree” configuration (Fukada 1997). This arrangement is shown in Figure 4.3.

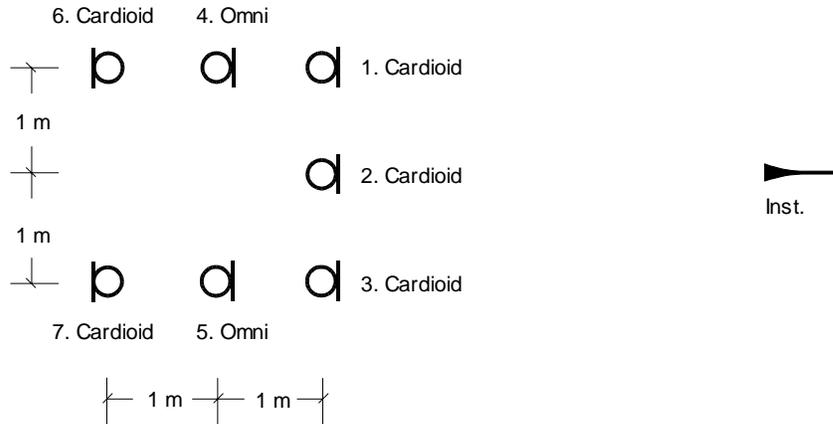


Figure 4.3 Microphone configuration used for the creation of impulse responses for measurements and listening tests. Note that the front-centre cardioid is located at the “centre” of the array. Not drawn to scale.

4.3 Creation of sample sound files for measurement and listening tests

Using this configuration, three individual multichannel impulse responses were created in MATLAB. The first of these contained only the direct sound, and thus was comprised of a simple delay with appropriate gain. The second multichannel impulse response used a perfectly specular reflective model for all four walls in the simulated concert hall. This resulted in the equivalent of a 3 to 4 tap delay, depending on the location of the microphone. (Since the front centre microphone and the instrument were both positioned in the exact centre of the hall, the reflections from the two longer walls arrive simultaneously and thus result in a total of 3 rather than 4 taps as in the case of all other microphone outputs.) The third and final multichannel impulse response used a perfectly diffuse model for all four walls with the design parameters listed in Table 4.2.

| Wall | Design Frequency | N |
|-------|------------------|----|
| North | 550 Hz | 17 |
| South | 750 Hz | 17 |
| East | 1100 Hz | 17 |
| West | 1300 Hz | 17 |

Table 4.2 Schroeder diffuser parameters used in the diffuse reflection model impulse response.

4.3 Creation of sample sound files for measurement and listening tests

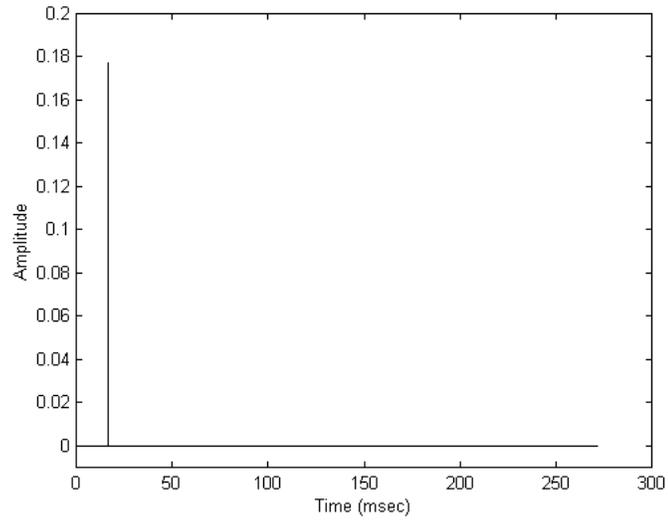


Figure 4.4 Impulse response of the direct sound from the left omnidirectional microphone.

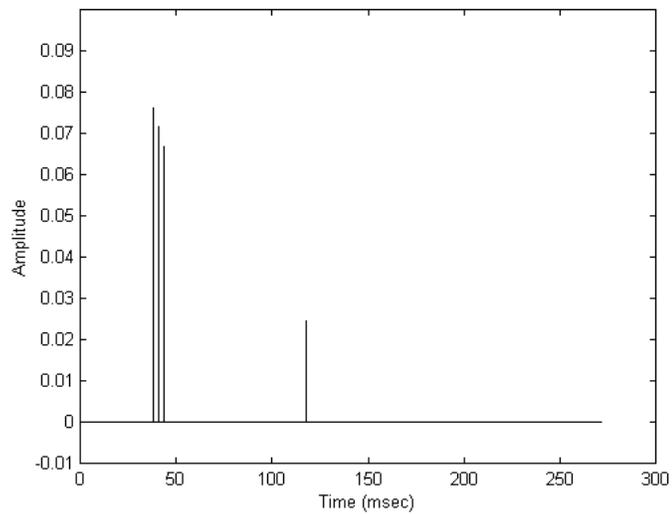


Figure 4.5 Impulse response of the specular reflections from the left omnidirectional microphone.

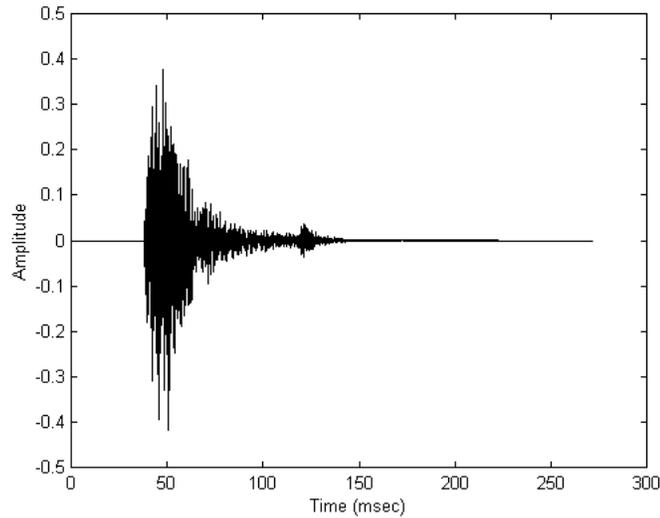


Figure 4.6 Impulse response of the diffused reflections from the left omnidirectional microphone.

The resulting nine sound files (three original sound files each convolved with three impulse responses) were exported from MATLAB in the Sun “.au” format and converted to a standard Audio Interchange File Format (.aiff) file using the SoundHack shareware for Macintosh. This procedure was used since MATLAB is unable to export the required .aiff files directly. These files were subsequently imported into a custom patcher created in the Max/MSP software package for Macintosh. This software is chosen for its simplicity of use in creating custom DSP patches as well as its internal 32-bit resolution, necessary since MSP is used for some minor post-processing. This patcher permitted the synchronous playback of the direct, specular and diffuse reflection models and thus permitted the end user to adjust manually the relative balances in real time during playback.

4.4 Frequency response analysis

A prime requisite of the system is to produce a method of diffusing early reflections in order that they have a beneficial aesthetic effect on the program material. One principal method of analysis of the impulse response which can be used to predict this effect is a simple frequency response measurement. The analysis presented here is not a measurement but a calculation, using the impulse responses themselves.

Due to the differences in the characteristic temporal responses of the two reflection models, they have substantially different frequency response characteristics throughout their durations.

4.4.1 Entire impulse response

Frequency responses of the entire impulse responses for both the specular and diffused reflection models display some interesting characteristics. Figure 4.7 shows a normalized one-third octave smoothed frequency response calculated from a 65,536-point power spectral density (PSD) analysis in MATLAB. Although the plot indicates a relatively flat frequency response, $+2.5/-4.5$ dB from 20 Hz – 20 kHz, this is in fact deceiving. As can be seen in Figure 4.8, the same plot without the smoothing function displays the characteristic periodic curve of a comb filter, more evident in the linear frequency plot in Figure 4.9. Although this particular situation results in multiple resonant frequencies and harmonics, a result of the 4 offset impulses, it is clearly audible, particularly in low frequency bands.

4.4.1 Entire impulse response

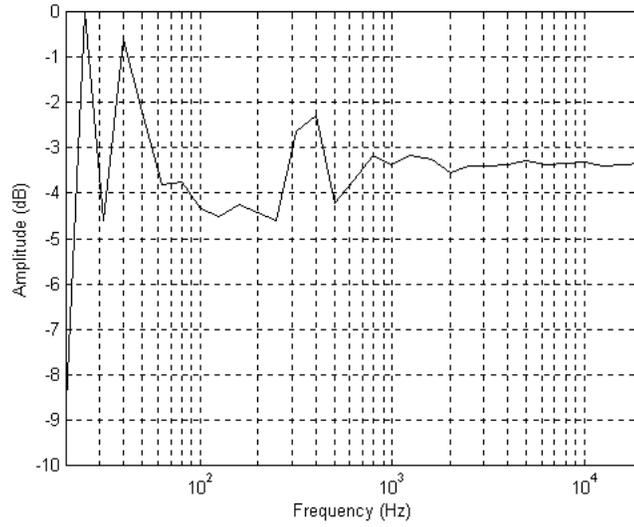


Figure 4.7 One-third octave smoothed normalized DFT of direct sound with specular reflections

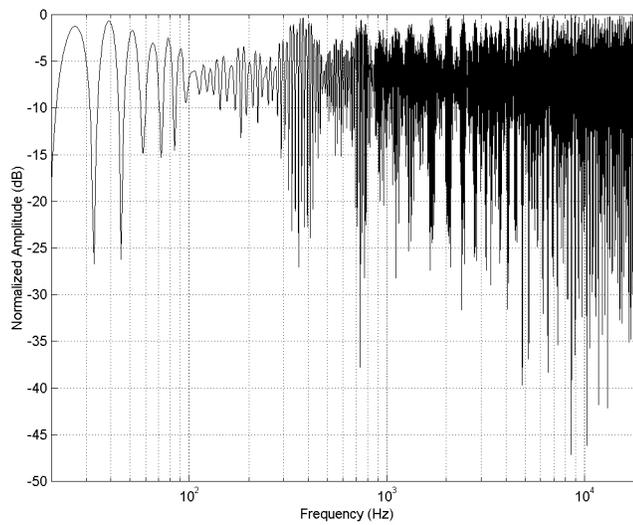


Figure 4.8 Normalized DFT of direct sound with specular reflections

4.4.1 Entire impulse response

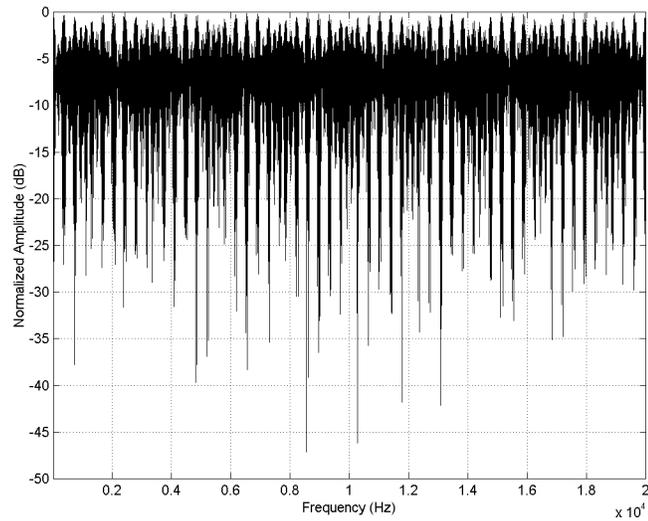


Figure 4.9 Normalized DFT of direct sound with specular reflections

In comparison, the normalized one-third octave frequency response of the diffused reflection impulse response in Figure 4.10 is far less flat, with a total range of approximately 40 dB from 20 Hz to 20 kHz. There is a noticeable boost of mid to high-mid frequency information caused by the resonances in the diffuser wells with roll-offs in the low and high frequency ranges. The complete plot in Figure 4.11 shows two characteristics. Firstly, there is a minor low-frequency comb filtering effect with a fundamental resonant peak below 30 Hz. Secondly, the periodicity of the resonances evident in the specular reflection plot is eliminated.

4.4.1 Entire impulse response

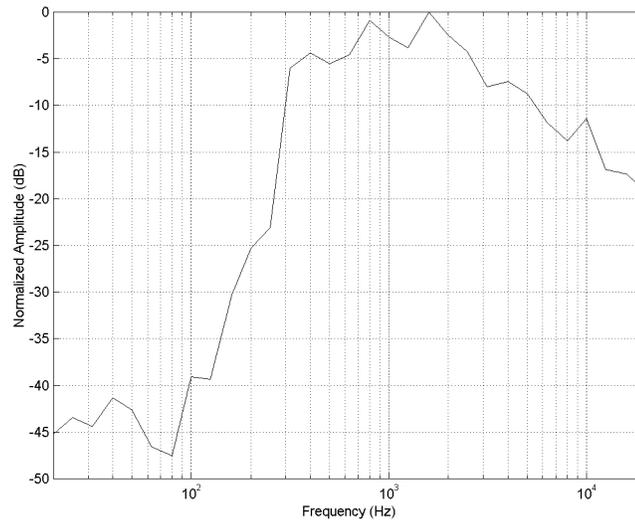


Figure 4.10 One-third octave smoothed normalized DFT of direct sound with diffused reflections

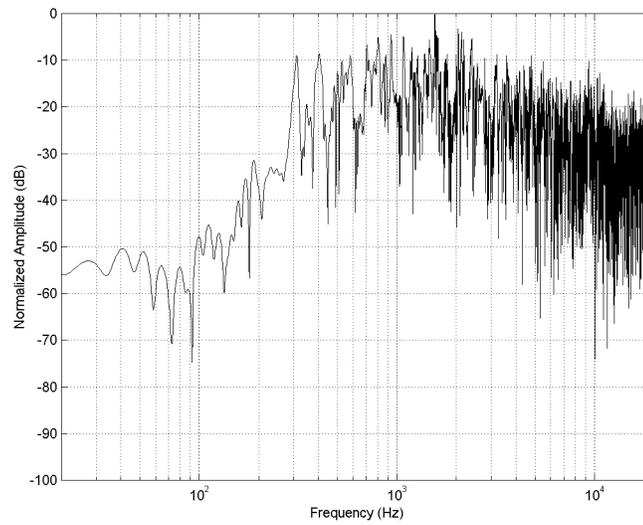


Figure 4.11 Normalized DFT of direct sound with diffused reflections

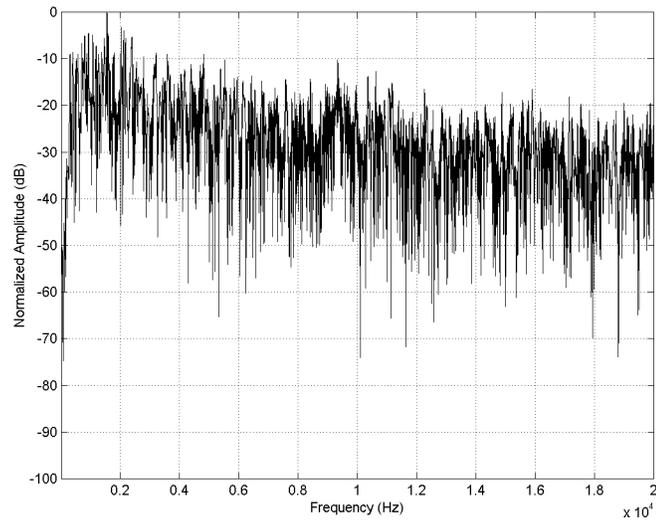


Figure 4.12 Normalized DFT of direct sound with diffused reflections

These frequency responses provide an overall concept of the timbre of each of the impulse responses. However they do not provide an indication of the change in frequency response over time. This is achieved by comparing both the waterfall and spectrogram plots of the same data.

4.4.2 Waterfall plots

A waterfall plot displays a number of frequency response graphs representing the change in the relative levels of various frequency bands over time. In this instance, this is achieved by calculating a power spectrum density for a subset of samples from the entire impulse response, storing it and continuing to the next subset. The waterfall plots shown in Figure 4.13 and 4.14 display the square roots of one-third octave 65,536-point PSD's

4.4.2 Waterfall plots

of subsets of 1000 samples, taken every 1000 samples (i.e samples 1-1000, 1001-2000, ...). This is equivalent to a frequency response calculation every 22.7 ms.

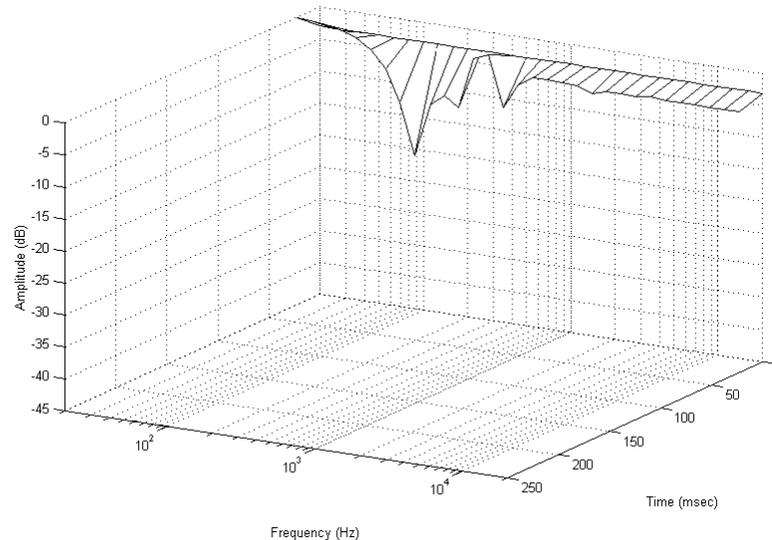


Figure 4.13 One-third octave smoothed normalized waterfall plot of direct sound with specular reflections

Figure 4.13 shows the one-third octave smoothed waterfall plot for the direct sounds and specular reflections. There are two basic principal characteristics worthy of discussion. Firstly, the flat frequency response of the direct sound is visible at time 0. This is due to the fact that the earliest reflection occurs later than 1000 samples after the beginning of the impulse response, therefore the only content of the first subset is a single gain-reduced impulse. The second is the apparent lack of any information following the second subset. This is in fact, a problem in plotting rather than a representation of the signal. As can be seen in Figure 4.5, the temporal response of the specular reflections following the 2000th

sample consists only of a single sample reflection at sample number 5196. Although the FFT of this reflection has a flat frequency response at a level of approximately -43 dB relative to the 0 dB in the waterfall plot in Figure 4.13, it is preceded by three subsets of FFT's with values of $-\infty$ dB.

Also of note are the characteristics of the frequency response of the second subset. This is effectively a frequency response measurement of the three earliest specular reflections in isolation from the direct sound. There are two main components of this curve: the first is a boost in the extremely low frequencies due to phase correlation. This results in a total level greater than the direct sound. The second is a more “traditional” comb filter curve which is due to the closely-matched amplitudes and the nearly-regular spacing of the three earliest first reflections.

4.4.2 Waterfall plots

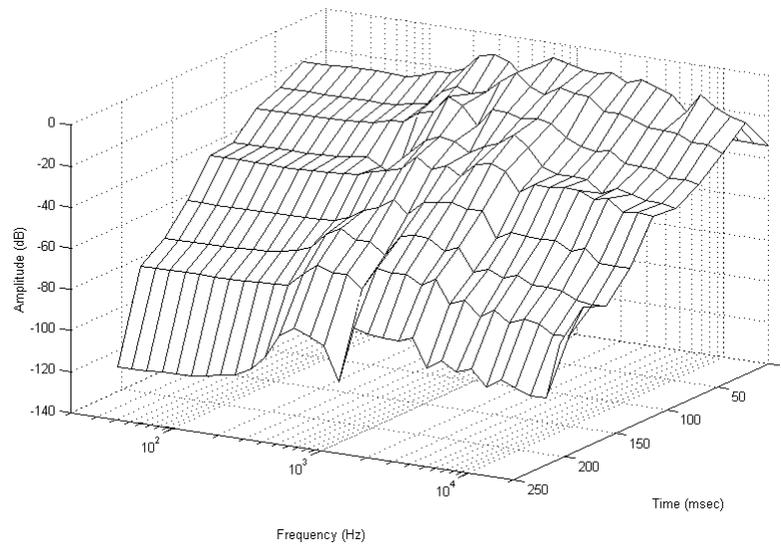


Figure 4.14 One-third octave smoothed normalized waterfall plot of direct sound with diffused reflections

The waterfall plot of the purely diffuse reflections in Figure 4.14 shows a considerable difference from its specular counterpart. The decay times of all frequency bands are considerably longer, lasting a total of roughly 12,000 samples (approximately one quarter of a second) to decay approximately 100 dB. Although not evident in the displayed plot, similar to the specular reflections, the first subset of 1000 samples has a flat frequency response since it consists of only the direct sound. Note that the amplitude of the boost in the high midrange displayed in the frequency response of the complete impulse response in Figure 4.14 is considerably reduced to a window of approximately 20 dB in the waterfall plot.

4.4.3 Spectrograms

The waterfall plots displayed above show an averaged frequency response of a number of large sections of the temporal response. In order to have a more detailed view of the progression of resonant frequencies in time, we must also use a spectrogram plot of the impulse responses.

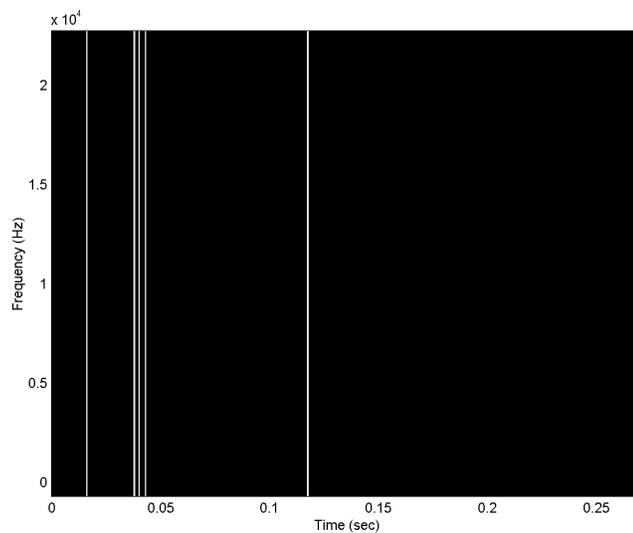


Figure 4.15 Spectrogram of the impulse response of the specular reflection model. Note the wideband frequency response of each of the four reflections in the impulse response. Brighter points indicate more power at the corresponding time and frequency.

Figure 4.15 shows a spectrogram of the impulse response of the specular reflection model. Four bright vertical lines are visible in an otherwise black graph. These four lines correspond to the four individual reflections, each a gain-reduced replica of the original

4.4.3 Spectrograms

Dirac impulse. Since a Dirac contains a flat frequency response from DC to the Nyquist frequency, each of these lines displays a single intensity for all frequencies. This plot was calculated in MATLAB using an FFT size of 32 points in order to optimize the time resolution of the system. Although this results in a very poor frequency resolution, this is insignificant since each reflection has a predictably flat frequency response from DC to the Nyquist frequency.

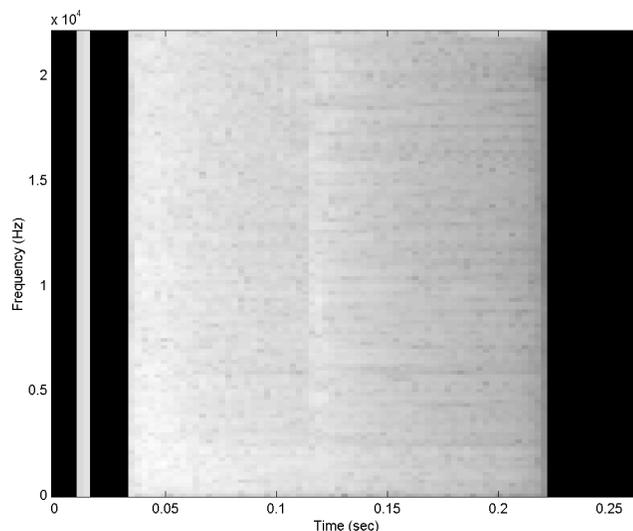


Figure 4.16 Spectrogram of the impulse response of the diffuse reflection model. Brighter points indicate more power at the corresponding time and frequency.

Figure 4.16 shows the spectrogram using a 256-point FFT of the diffuse reflection model. A number of characteristics previously unseen are evident in this plot. The first is the delineation between the first three reflections and the fourth from the more distant wall in the room model at approximately 0.12 sec. The second is the relatively even distribution

of power throughout the decay. Also, note that there is a lack of comb filtering in the plot, recognizable in a spectrogram as a series of regularly-spaced horizontal lines.

4.5 Comparative cross-correlation measurements

One of the reasons behind the initial development of the Schroeder diffuser for real spaces was to decrease the level of interaural cross correlation at various listening positions for the audience members. In order to determine the response of this synthesized implementation in this regard, an IACC value must be measured rather than calculated. This is due not only to the fact that the system is intended for playback over a multichannel audio reproduction system but that an IACC measurement must necessarily incorporate the *head related transfer functions* (HRTF's) of a human head. Consequently, a calculation of the IACC based on the manufactured impulse response will not necessarily correspond to values at the listening position in a real monitoring environment.

4.5.1 Measurement and reproduction equipment

In order to make the IACC measurements, a Brüel & Kjær Head and Torso Simulator (HTS) Type 4100 (Serial number 1665903) was placed at the listening position behind the computer monitor in the Multichannel Audio Research Laboratory (See Appendix C for a floorplan and elevation of the MARLab). The height of the ear canals of the HTS was 108 cm (Note that this is slightly lower than the loudspeaker tweeter height of 155.6 cm) and the computer monitor had a maximum height of 103 cm. The outputs of the HTS were sent to a Brüel & Kjær two-channel Signal Analyzer Unit Type 2035 (Serial number 1715555) equipped with 100 kHz Input Module Type 3020. These input modules provide

both the 200 V DC voltage supply for the HTS and the necessary preamplification stage for the signal analyzer.

The signal analyzer was configured to measure the cross correlation of the real component of the two input channels with a linear average of 1000 spectra. The system was free-running (rather than triggered), used a Hanning weighting, recommended for “general purpose analysis of continuous signals” (Brüel & Kjær 1994) and was bandlimited from 22.4 Hz to 25.6 kHz. Although the measurements were performed early on a Saturday morning to minimize interference of low-frequency signals from external traffic noises, the 22.4 Hz high pass filter was engaged in order to further reduce this problem.

The reproduction system used the analog outputs of a Mark of the Unicorn (MotU) 2408 audio I/O device, providing five 16-bit 44.1 kHz outputs which were in turn connected to the analog line inputs of a Yamaha 03D digital mixing console. The outputs of the Yamaha were subsequently sent directly to the inputs of 5 Bang & Olufsen Beolab 4000 self-powered two-way loudspeakers arranged in a standard 5-channel configuration in accordance with the ITU-R BS.775-1 (ITU 1994) specification and as shown in Appendix C. Appendix D shows frequency response measurements of the five output channels including the loudspeakers.

The down mixing of the seven audio channels in each sound file to the five output channels was accomplished in the Max/MSP patcher. Following the Fukada Tree configuration, the output of each of the cardioid microphones was routed to its

corresponding single output channel (left-front microphone to left-front loudspeaker and so on...). Each of the two omnidirectional outputs was routed using equal gain to two loudspeakers albeit at a level 3 dB lower than the cardioids (left microphone to left-front and left-rear loudspeakers; right side symmetrical).

For the purposes of the IACC measurements, only the sound files with the calculated reflections were used. This was due to the fact that, as will be discussed in the following chapter, while the direct sound and a synthetic reverberation were used for the listening test, the interest in the IACC measurements was in the comparison of the two reflection models, and thus could be measured in isolation. These sound files were looped, providing continuous playback throughout the entire averaging process of the Signal Analyzer Unit.

Table 4.3 shows the measured sound pressure levels of the signals used in the IACC measurements. These measurements were made at the listening position using a Brüel & Kjær Precision Sound Level Meter Type 2235 (Serial number 1630394) equipped with the Type 4176 Microphone Capsule (Serial number 1651661), using a random incidence setting with slow averaging. Table 4.3 lists the range of the levels throughout each sound sample.

4.5.1 Measurement and reproduction equipment

| Sound file | Specular | Diffuse |
|------------|-------------|-------------|
| Xylophone | 61.0 – 71.8 | 64.6 – 71.1 |
| Speech | 59.9 – 66.7 | 55.1 – 59.4 |
| Cello | 61.3 – 83.9 | 63.0 – 75.7 |

Table 4.3 Sound pressure level measurements of the sound files used for IACC measurements. All measurements are in dB_{spl} A-weighted

4.5.2 Measurement environment characteristics

The ambient noise level in the MARLab was measured in advance of and following the IACC measurements using a Brüel & Kjær Precision Sound Level Meter Type 2235 (Serial number 1630394) equipped with the Type 4176 Microphone Capsule (Serial number 1651661). Using a Random incidence setting with slow averaging, the average noise floor of the test room was found to be an acceptable 26.6 dB_{spl} A-weighted at the listening position.

The RT₆₀ reverberation time of the monitoring environment was measured using a DRA Laboratories MLSSA and found to be 0.1 second at 1 kHz. Appendix C contains more detailed information both on the measurement procedure and on the results of the acoustical characteristics of the monitoring environment.

The reproduction chain was calibrated to a tolerance of approximately ± 0.5 dB from 20 Hz to 20 kHz. More thorough information regarding the frequency and time response of the electrical and electroacoustic components in the reproduction chain is detailed in Appendix D.

4.5.3 Results

Table 4.4 shows the results of the comparative IACC measurements of the three sound files. A number of interesting characteristics are revealed by these measurements. Firstly, it is evident that, in all cases, the diffuse model provides significantly reduced interaural cross correlation than for the specular model for all sound files. Secondly, it is interesting to note that the variation in IACC between the three sound files is smaller for the diffuse model than the specular model. This is particularly noticeable in the remarkable difference in the specular model between the xylophone sound file and the other two. Thirdly, while the specular model for the xylophone sound file produces decreased IACC measurements in comparison to the cello and speech files, the reverse is true for the diffuse model, although on a much smaller scale. This effect can be ignored as it is largely the product of the lack of low-frequency content in the sound sample. As a result, the IACC is lowered for the specular model due to very small differences in the channel outputs and slight inaccuracies in the placement of the HTS. Since the diffuse model results in an averaged signal due to time smearing of the impulse response, it is less affected by these small errors. One possible conclusion of this is that the optimal listening position or so-called “sweet spot” is enlarged by the use of the diffuse model, however this is speculative, based solely on these measurements and then, pertinent only to high frequencies.

4.5.3 Results

| Sound file | Specular | Diffuse |
|-------------------|-----------------|----------------|
| Xylophone | 0.118 | 0.076 |
| Speech | 0.435 | 0.061 |
| Cello | 0.449 | 0.062 |

Table 4.4 Interaural cross correlation for three sound files with the HTS at the optimal listening position.

In order to investigate further this hypothesis, additional IACC measurements were performed with the HTS placed at locations other than the recommended listening position. Three additional locations, 25 cm, 50 cm and 100 cm to the left of the optimal “sweet spot” were chosen for measurements. Note that the 100 cm location is an extreme case in that, at this location, the HTS centre is approximately 283 cm from the right-rear loudspeaker but only 132 cm from the left-rear loudspeaker. The results of these measurements are listed in Tables 4.5 for the xylophone sound sample, and 4.6 for the speech.

| | 0 cm | 25 cm | 50 cm | 100 cm |
|-----------------|-------------|--------------|--------------|---------------|
| Specular | 0.118 | 0.042 | -0.178 | 0.264 |
| Diffuse | 0.076 | 0.098 | -0.236 | 0.019 |

Table 4.5 IACC measurements with the HTS situated at various distances from the optimal listening location for the xylophone sound sample. Distances indicated are directly left of the “sweet spot”

4.5.3 Results

| | 0 cm | 25 cm | 50 cm | 100 cm |
|-----------------|-------------|--------------|--------------|---------------|
| Specular | 0.435 | 0.576 | 0.472 | 0.740 |
| Diffuse | 0.061 | 0.041 | 0.011 | 0.291 |

Table 4.6 IACC measurements with the HTS situated at various distances from the optimal listening location for the speech sound sample. Distances indicated are directly left of the “sweet spot”

Regrettably, the results of the measurement are inconclusive. Although informal listening tests suggest that the diffuse reflection model provides an impression of an enveloping environment in a larger area surrounding the optimal listening location, this sensation is not supported by the IACC measurements. In the cases of both xylophone and speech sound samples, there is no notable trend in the cross correlation values with movements of the listening position for either reflection model.

4.6 Imaging characteristics

One important characteristic of artificial reverberation systems, particularly of the early reflection response, is the spatial placement of the various components of the signal. Regardless of the number of output channels, in order to achieve a controllable sense of width and depth, the various reflections must be appropriately distributed in the listening space. It is common practice for manufacturers to follow the philosophy used in the second implementation of the SceneBuilder package. The intention of this topology is similar to that used by Moorer (1979) in that it attempts to convert the listening space into an alternate acoustic environment by locating the various components of the spatial sound in expected locations relative to the listener in the real space. This is accomplished differently for the late and early components of the reverberation.

It is typically assumed that the late reverberant energy in synthesized acoustic spaces is equivalent to a diffuse field due to the high number of reflections arriving at the listening position from all directions. As a result, when determining the imaging characteristics of the reverberant tail in a synthetic acoustic environment, manufacturers and programmers usually concern themselves only with the inter-channel correlation values. The traditional methodology is to create as many uncorrelated reverb tails as there are output channels which are cross-mixed to produce the desired correlation.

In contrast, the early reflections are individually panned to various locations across the sound stage. Depending on the particular device, these locations are either preset by the manufacturer or can be controlled by the end user. Various industry representatives have

espoused the use of simple pair-wise power panning to achieve this steering for various reasons, principally improved localization accuracy and monophonic compatibility. Although these reasons are quite valid, the conclusion is not necessarily preferable to alternate panning techniques (Martin 1999a).

As was discussed in Chapter 1, research in the MARLab has indicated that the use of virtual microphones with various locations in the synthetic enclosure provides a greater sensation of space and realism for the virtual acoustic environment. As in a real recording situation, the resulting imaging characteristics are consequently heavily dependent not only on the location of the instrument and the characteristics of the recording space, but on the locations and polar patterns of the various microphones (Martin 1996).

In its present implementation, the system described herein is intentionally designed to mimic the characteristics of microphone polar patterns. However, with modification, it can be easily converted to emulate other panning algorithms. There are two significant issues to discuss regarding the implementation of this emulation: the time and gain response of the various output channels of the system. The former of these two can be aligned simply by arranging all microphone locations in the system to the same position, creating a coincident configuration. Time differences between the various output channels will be thereby automatically eliminated, thus providing panning based only on amplitude differences between the loudspeakers.

4.6.1 Microphone implementation

The imaging characteristics of the system in its described implementation emulate those of a pair of microphones in a real space. Regrettably, little research and analysis exists in the area of imaging response characteristics of microphone arrays for multichannel audio reproduction. As a result, the perceptual merits of the system remain to be evaluated. The initial objective of using an algorithm centred on virtual microphones was to improve both the perceived audio quality and the user interface for the system. While preliminary informal tests in the MARLab indicate that the system excels in both respects, a systematic evaluation based on blind listening tests is required to make reliable conclusions in this regard.

4.6.2 Pair-wise power panning emulation

In order to convert the system so that it provides an output such as that generated by more common topologies, the polar patterns of the virtual microphones must be altered to generate simple pair-wise constant power panning similar to that produced by a pan knob on a recording console. To begin, the microphones must be coincidentally located. The system, described in Equations 4.2 and 4.3, is based on a gain determined by a function of the desired angle to the sound source ξ relative to the angle of the loudspeakers (Roads 1996).

$$G_L = \frac{\sqrt{2}}{2} [\cos(\xi) + \sin(\xi)] \quad (4.2)$$

4.6.2 Pair-wise power panning emulation

$$G_R = \frac{\sqrt{2}}{2} [\cos(\xi) - \sin(\xi)] \quad (4.3)$$

where ξ is the desired angle to the phantom image between the loudspeakers and loudspeakers A and B are assumed to be located at $\xi = \pm 45^\circ$. See Figure 4.17 for a plot of these responses.

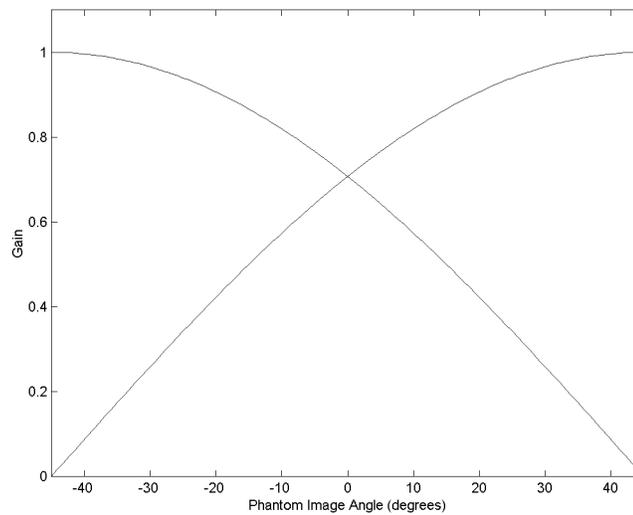


Figure 4.17 Gain curves for pair-wise power panning for two loudspeakers located at $\pm 45^\circ$.

This gain function would subsequently have to be modified to accommodate the configuration using 5 loudspeakers with irregular angular spacing. Figure 4.18 shows the gain plots of this function customized for loudspeakers at 0° , $\pm 30^\circ$, and $\pm 120^\circ$.

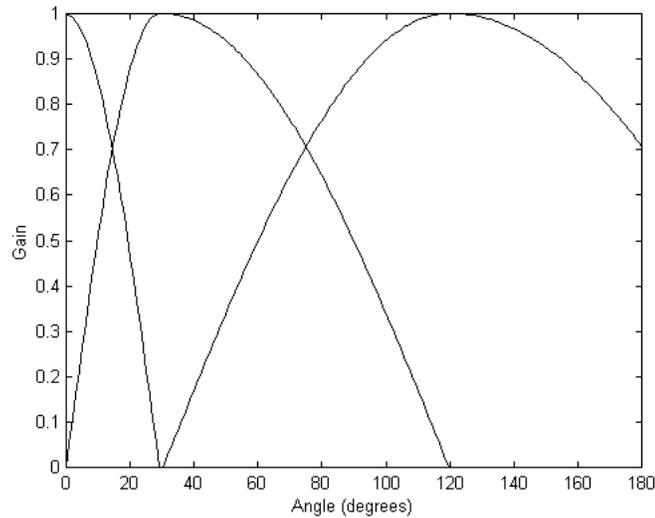


Figure 4.18 Power panning gain functions for a five-loudspeaker configuration. (From left to right, Centre, Right and Right Surround curves). Only the right side is shown as the left side is symmetrical.

There are two principal issues to note when considering this system as a modification to the virtual microphone implementation described above. Firstly, if the loudspeaker configuration uses an irregular angular spacing, then the calculated “polar patterns” of the microphones would necessarily be irregular as demonstrated in Figure 4.18. For example, note the asymmetry of the right and right surround panning curves. Also, there is a large angular area in each gain function in which the gain of the system is 0. Secondly, this function can allow for multiple output channels and therefore pair-wise constant power panning over any number of distributed loudspeakers. There must, however, be one virtual microphone for each loudspeaker with a customized gain vs. angle function.

4.6.3 Ambisonics emulation

The system can be used without modification to generate quite easily a three-channel first-order pantaphonic Ambisonics B-format output. This is due to the fact that the format is equivalent to the outputs of a coincident cluster of three microphones, one omnidirectional (the so-called “W” channel) and two bidirectional, arranged at 0 degrees and 90 degrees of rotation (the “X” and “Y” channels respectively).

$$W_a = P_a \tag{4.4}$$

$$X_a = P_a \cos(\xi_a) \tag{4.5}$$

$$Y_a = P_a \sin(\xi_a) \tag{4.6}$$

where W_a , X_a and Y_a are the three channels in the Ambisonics B-format output corresponding in the real world to omnidirectional, front-facing bidirectional and side-facing bidirectional microphones respectively, P_a is the incoming pressure and ξ_a is the angle to sound source a at the microphone position.

The resulting audio signals would then be converted through a standard B-format decoder to output to any number of speakers in any locations using the formula shown in Equation 4.7 (Bamford 1995).

$$G_b = \frac{W_a + 2X_a \cos(\Psi_b) + 2Y_a \sin(\Psi_b)}{B} \tag{4.7}$$

where G_b is the gain of loudspeaker b , Ψ_b is the angular location of loudspeaker b and B is the total number of loudspeakers.

The objective of the Ambisonics system is to replicate the wavefront characteristics of the recorded space in the listening space. The accuracy of this reproduction is dependent on the order of the system being employed, with various components of the wavefront included with various orders. The zeroth-order Ambisonics system reproduces the pressure component of the wavefront and employs only the W_a -channel – the zeroth order spherical harmonic – in the B-format signal. The inclusion of first-order spherical harmonics – the X_a and Y_a channels in the first-order Ambisonics system – reproduces the velocity, and therefore propagation direction, of the wavefront. Augmenting the system with second-order spherical harmonics, based on the cosine of twice the angle of incidence of the incoming soundwave provides the listener with the curvature of the wavefront.

Although it is not possible to achieve this second-order system with physical microphones, it can be realized in the theoretical model described here by altering once again the polar patterns of the microphones. In order to produce the second-order spherical harmonics, the polar patterns described in Equations 4.8 and 4.9 are used (Bamford 1995).

$$U_a = P_a \cos(2\xi_a) \tag{4.8}$$

$$V_a = P_a \sin(2\xi_a) \tag{4.9}$$

where U and V are the two additional channels in the Ambisonics B-format output and P_a and ξ_a are the incoming pressure from and angle to sound source a at the microphone position (Bamford 1995).

$$G_b = \frac{W_a + 2X_a \cos(\Psi_b) + 2Y_a \sin(\Psi_b) + 2U_a \cos(2\Psi_b) + 2V_a \sin(2\Psi_b)}{B} \quad (4.10)$$

Where G_b is the gain of loudspeaker b , Ψ_b is the angular location of loudspeaker b and B is the total number of loudspeakers.

Higher degrees of accuracy are possible with higher orders of the Ambisonics channels, however the expense is in the form of loudspeakers. In the case of pantaphonic systems, the minimum number of loudspeakers required to accurately reproduce the wavefront is computed using Equation 4.11 (Bamford 1995).

$$B_{min} = 2m + 1 \quad (4.11)$$

where B_{min} is the minimum number of loudspeakers required and m is the order of the pantaphonic Ambisonics system.

4.6.4 Binaural simulation / Auralization

It is possible to adapt the system to provide a binaural output for headphone reproduction if the virtual microphone component is replaced by a filtering system based on HRTF's. The angle of incidence calculated from the location of the point of reflection for each sample in the impulse response would determine the appropriate HRTF measurement to

be used. The result of the well time response would then be convolved through the HRTF before being added to the complete impulse response. Such a system would be similar to the existing auralization software packages and could be implemented as a module applied to the system for processing first reflections .

5 Listening tests

Although some characteristics of the system can be analyzed using mathematical computation and electroacoustic measurements, such an analysis would not necessarily constitute an evaluation of the preferability of the procedure as a method of processing audio signals. Such an evaluation must be conducted by means of listening tests performed by human listeners who are asked to indicate their preferences when presented with various models of synthetic early reflections. Since almost all systems currently in use in recording environments make use of a specular model for these reflections, this will be used as the reference model for the comparison tests.

The first step in the development of a methodology for psychoacoustic evaluation is the determination of the question to be answered by the investigation. In this case, the evaluation process seeks to determine whether the system described in this dissertation is preferred by listeners to the traditionally-used specular model of early reflections.

The evaluation process consisted of two rounds of formal listening tests with distinctly different objectives, conducted on two separate occasions.

The first round of tests sought to evaluate the ability of listeners to distinguish between reflection models based on completely specular or completely diffused distributions of energy.

The second round determined the preferences of listeners presented with the ability to mix the relative levels of the two models. If the test subjects in the first round of listening

tests were unable to detect a difference between completely opposite models, then the proposed system would be effectively irrelevant and it would be futile to proceed to the second round of tests. If, however, the system is shown to be detectable, then the second tests would be executed to determine if the system is preferred by listeners.

5.1 Technical descriptions

The system used for all listening tests is a standard hardware configuration used in the MARLab. The entire system including software and hardware, as well as electroacoustic and acoustical properties of the room itself are considered in order to ensure reliable test data.

5.1.1 Hardware and software

The software platform used to create the listening tests was Cycling '74's "Max/MSP." This is a graphics-based programming environment for the creation of real-time DSP processes running on an Apple Macintosh. All internal calculations are done at 32 bits floating point of resolution. The input and output hardware consisted of a Mark of the Unicorn 2408 which communicates with the Macintosh using a proprietary digital audio protocol and ASIO drivers. The analog outputs of the 2408 were connected to the analog inputs of a Yamaha 03D digital mixer which was used for channel level calibration. The outputs of the Yamaha were connected to 5 Bang & Olufsen Beolab 4000 self-powered two-way loudspeakers arranged in accordance with the ITU-R BS.775-1 specification (ITU 1994).

5.1.1 Hardware and software

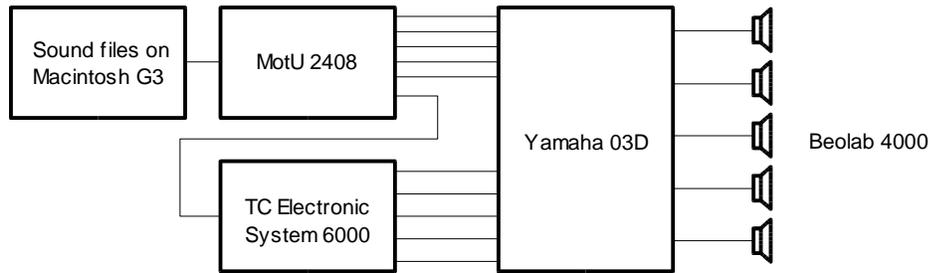


Figure 5.1 Block diagram of hardware configuration

A number of sound signals included a reverberant component provided by a t.c. electronic System 6000 multichannel digital reverberation processor in real time during the tests. This is a reverberation unit intended for 5.1-channel output, providing five discrete uncorrelated reverberation tails. One input of this device was connected to an analog output of the 2408 and its digital outputs were connected to the AES/EBU inputs of the Yamaha 03D. The sampling rate of the device was 44.1 kHz. It provided the master clock source for the mixing console. The parameters of the Engine used on the reverberation unit for all listening tests are listed below.

Software versions

| | |
|----------------|---------|
| Frame Software | v. 1.25 |
| DSP Software | v. 1.25 |

Engine E1 - Main

| | |
|--|----------------|
| Name | Concert Hall |
| Algorithm | VSS 5.1 Source |
| All parameters set to default or 0.0 dB unless otherwise listed. | |
| Master decay | 1.5 seconds |
| Reverb size | 15 |

Setup

| | |
|----------------|-----------|
| Output Mode | 5.1 Music |
| Centre Channel | On |
| Dry Signal | Off |

View

| | |
|---------------|--------------|
| Location Type | Concert Hall |
| Source 1 | C 0° |

Sources

| | |
|----------|-------|
| ER Start | 0 % |
| ER Stop | 100 % |

Source 1 Levels

| | |
|--------------|----------|
| ER Level | Off |
| Reverb Level | 0.0 dB |
| Dry Level | Off |
| In Level | -30.0 dB |

Reverb

| | |
|-----------------------|------|
| Mid Color | 50 |
| All decay multipliers | 1.00 |

Modulation

| | |
|-----------------|-----|
| Modulation Type | Off |
|-----------------|-----|

The parameters of the reverberation device were arbitrary values chosen almost entirely on the basis of aesthetic considerations. They were, however, intended to match roughly the reverberant characteristics of McGill's Redpath Hall, the room size used to model the first reflection pattern. The various signals with either specular or diffuse reflections were played simultaneously with the reverberation and the input level was adjusted to achieve

a desirable balance. As discussed in Section 5.1.3, this value was measured to be approximately 17 dB below the direct level. The measurement, however, was made only for documentation purposes.

5.1.2 Sound sources

In order to determine if listener responses would differ according to varying types of signals, three different sound sources were selected. Each of these was chosen to represent a distinct aspect of sound signal characteristics. All recordings are anechoic samples from the Bang & Olufsen test disc *Music for Archimedes* (Bang & Olufsen 1992). The first sample is an excerpt of solo xylophonist playing “Sabre Dance” by Khachaturian. This was chosen to highlight the transient characteristics and high-frequency response of the system. The second sample is an excerpt of Weber’s “Variation and Theme No. 2” played on a cello. This was chosen to highlight both low-frequency and steady-state responses of the system. The third is a recording of female speech, used as a combination of transient and steady state, and because speech is frequently used in listening tests as an equivalent and less tiring replacement for noise-based signals (Kleiner 1993). The test CD includes examples of speech in both English and Danish – the latter was chosen to avoid distraction.

Technical information regarding the recording techniques and equipment for the sound samples used can be found in Appendix B.

For each stimulus in each test, the subject was randomly presented with and asked to compare two signals, one with specular reflections and one with diffused reflections, from one of the three sound sources. For a given stimulus, the signal might or might not include a reverberant tail. This option was included to determine whether subjects' responses changed with the inclusion of reverberation. Further details of the individual tests are outlined in Sections 5.2 and 5.3.

5.1.3 Equipment configuration and alignment

All listening tests were conducted in the Multichannel Audio Research Laboratory at McGill University's Redpath Hall. This is a room specifically designed for multichannel listening tests and research and is equipped with extensive acoustical treatment. Details of the acoustical properties of the room can be found in Appendix C while measured data of the reproduction equipment are listed in Appendix D.

The convolution of a sound file through the impulse responses of the specular and diffused models results in significant level differences. In order to evaluate accurately the various models in psychoacoustic listening tests, it must be ensured that more obvious cues such as level differences are not considered by the listening subjects in making their decisions. Consequently, the playback levels of the two models were aligned to minimize audible differences. This was achieved by convolving a sample of pink noise through the same impulse responses used for the processing of the audio files. The resulting signals were then played through the multichannel reproduction system using the same parameters as the three anechoic recordings. The output of the system was measured

5.1.3 Equipment configuration and alignment

using a Brüel & Kjær Precision Sound Level Meter Type 2235 (Serial number 1630394) equipped with the Type 4176 Microphone Capsule (Serial number 1651661) placed vertically at the listening position using a random incidence setting. Several combinations of direct, reflected and reverberant sound were tested to verify the matched signal levels. The results of these measurements are listed in Table 5.1. Note that, in order to ensure a similar output level between the specular and diffuse reflections, a gain of 0.21 was applied to all diffuse output channels. This value was used as a static gain for all listening tests.

| Signal (Pink noise) | Level dB _{spl} A |
|---|---------------------------|
| Direct | 57.3 |
| Specular reflections only | 52.2 |
| Diffuse reflections only | 52.2 |
| Reverberation only | 40.1 |
| Direct + Specular reflections | 58.5 |
| Direct + Diffuse reflections | 58.5 |
| Direct + Specular reflections + Reverberation | 58.6 |
| Direct + Diffuse reflections + Reverberation | 58.6 |

Table 5.1 Output levels of the various components of a signal consisting of pink noise convolved through the same impulse responses as the test signals. Values are A-weighted with slow averaging.

5.1.4 Test subjects

The subjects engaged for the listening tests are all students and instructors from the McGill University program in sound recording and volunteered to participate. All were presented with a small remunerative gift following the test.

5.1.4 Test subjects

For the first test, two females and nine males ranging in age from 24 to 49, with no stated hearing impairments, participated. The group consisted of seven undergraduate, two graduate-level and two doctoral students. All are practicing recording engineers experienced in critical listening on a daily basis with a technical knowledge of recording procedures and can therefore be considered to be an expert listening group (Stone and Sidel 1993).

Almost the same group was used for the second test with one additional graduate-level and one fewer doctoral students participating. This test was conducted one week after the first.

5.2 Test 1: A / B / X test

The first of the two listening tests was conducted in order to determine whether listeners were able to distinguish between the completely specular and completely diffused models. This was implemented as an A / B / X test in which the subjects were presented with a stimulus consisting of a reference signal labeled “X” and were asked to choose which of two test signals, labeled “A” and “B” was identical to the reference. Table 5.2 lists the 12 sound signals used for the reference signal “X.” The “A” and “B” test signals matched the “X” signal in all parameters except for the early reflection model. The software randomly assigned a model to each of the test signals for each stimulus.

| Stimulus | Sound file | Reverberation | ER Model |
|----------|------------|---------------|----------|
| 1 | Speech | No | Specular |
| 2 | Speech | No | Diffuse |
| 3 | Speech | Yes | Specular |
| 4 | Speech | Yes | Diffuse |
| 5 | Xylophone | No | Specular |
| 6 | Xylophone | No | Diffuse |
| 7 | Xylophone | Yes | Specular |
| 8 | Xylophone | Yes | Diffuse |
| 9 | Cello | No | Specular |
| 10 | Cello | No | Diffuse |
| 11 | Cello | Yes | Specular |
| 12 | Cello | Yes | Diffuse |

Table 5.2 List of stimuli used as the reference signal “X” in Test 1. Signals “A” and “B” in each stimulus were randomly assigned to the two Early Reflection models without changing other variables.

Each reference stimulus was presented six times, resulting in 72 stimuli for the total test. These stimuli were presented in random order for each subject. The average time taken to complete this test was less than 30 minutes.

All subjects underwent a training session one week before the test in which they responded to 72 stimuli. These sessions began with a set of standardized verbal instructions and a demonstration of the system using a training version of the software.

For an illustration of the following test description, please refer to the screen shot of the test shown in Figure 5.2. Each stimulus began immediately after the subject clicked on the “Next” button on the screen. The reference signal began playing immediately and looped for continuous playback. The two test signals were played synchronously with the reference signal and could be monitored individually by clicking on the “A” and “B” buttons displayed at the top of the screen or on the corresponding keys on the keypad. In order to avoid a noticeable click or “bump” when switching between different signals, a 50 msec crossfade was implemented. Immediately below the buttons were two displays, one indicating the signal being monitored at the time, the other displaying the subject’s answer. All corresponding data were stored in a tab-delimited text file when the subject moved to the following stimulus upon clicking the “Next” button. Participants were also given the option of using four corresponding keys on the keypad in the event that they wished to work with their eyes closed. Subjects were not given any clues as to the differences between the signals.

Appendix E shows a copy of the written instructions presented to subjects on a window of the computer monitor during the test.

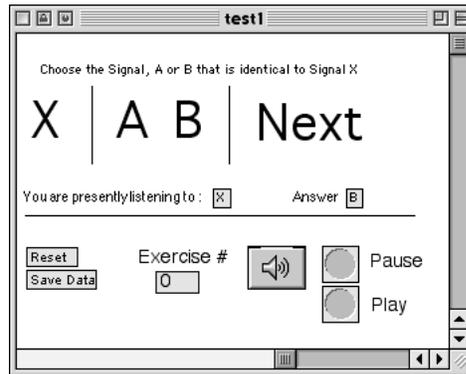


Figure 5.2 Screen shot of test window used in A / B / X test

Subjects were invited to rest during the test if they wished by clicking the “Pause” and “Play” buttons which controlled playback of the audio only.

5.2.1 Results

The results of the first test indicate that there is an easily recognizable difference between the specular and diffused reflection models. As shown in Table 5.3, of the twelve stimuli, one resulted in a score of 100% and ten resulted in accuracy over 90%. The lowest test scores resulted in an accuracy of 85%. The maximum standard error for the signals, a measurement of the variability of the associated results, was calculated to be 4%. This error could be reduced by increasing either the number of test subjects or the number of stimuli tested (or both), however, the results were adequate in all cases to provide a reliable indication.

5.2.1 Results

| Signal | Sound file | Reverb | ER Model | Result | Std. Error |
|--------|------------|--------|----------|--------|------------|
| 1 | Speech | No | Specular | 0.94 | ± 0.03 |
| 2 | Speech | No | Diffuse | 0.97 | ± 0.02 |
| 3 | Speech | Yes | Specular | 0.85 | ± 0.04 |
| 4 | Speech | Yes | Diffuse | 0.88 | ± 0.04 |
| 5 | Xylophone | No | Specular | 0.92 | ± 0.03 |
| 6 | Xylophone | No | Diffuse | 0.97 | ± 0.02 |
| 7 | Xylophone | Yes | Specular | 0.92 | ± 0.03 |
| 8 | Xylophone | Yes | Diffuse | 0.94 | ± 0.03 |
| 9 | Cello | No | Specular | 0.95 | ± 0.03 |
| 10 | Cello | No | Diffuse | 1.00 | ± 0.00 |
| 11 | Cello | Yes | Specular | 0.97 | ± 0.02 |
| 12 | Cello | Yes | Diffuse | 0.92 | ± 0.03 |

Table 5.3 Results of first listening test sorted by the twelve reference signals. The rightmost column indicates the number of correct answers for each stimulus.

χ^2 (chi-square) analysis was performed for various permutations of possible reference stimuli attributes to verify independence from the responses (Siegel 1996). This was accomplished by summing the results with similar attributes and comparing them with the summation of the results of the opposite attribute. Six tests were done for the following possibilities for the reference signal:

1. Reflection model: specular vs. diffused
2. Presence of reverberation: with reverb vs. without
3. Choice of sound file: All three sound files combined
4. Choice of sound file: speech vs. xylophone

5.2.1 Results

5. Choice of sound file: speech vs. cello

6. Choice of sound file: xylophone vs. cello

The results of these tests are listed in Table 5.4. P-values equal to or less than 0.05 indicate that there is a significant relationship between the combination tested and the results of the test. P-values equal to or less than 0.01 indicate that there is a very significant relationship between the two. For example, there is no significant relationship between the choice of specular or diffused early reflection model and the result of the test, therefore, either choice will produce similar results. However, since the P-value of the grouping by reverberation is 0.009, the decision to include reverberation in a stimulus will have very significant effect on the results of the test. As can be seen in Table 5.4, there is a significant relationship between the choice of sound file, in particular the speech vs. cello signals, and the results of the test. Consequently, the system must be tested with all sound files and reverberation combinations in the second listening test.

| Combination | χ^2 | df | N | P-value |
|--------------------------------|----------|----|-----|---------|
| Specular vs. Diffused | 1.366 | 1 | 792 | 0.242 |
| With vs. without reverberation | 6.917 | 1 | 792 | 0.009 |
| Choice of sound file | 6.319 | 2 | 792 | 0.042 |
| Speech vs. xylophone | 1.731 | 1 | 528 | 0.188 |
| Speech vs. cello | 6.162 | 1 | 528 | 0.013 |
| Xylophone vs. cello | 1.456 | 1 | 528 | 0.228 |

Table 5.4 Results of the chi-square analyses of the various combinations of parameter permutations (df = degrees of freedom).

5.2.1 Results

Listeners were invited to comment informally on the characteristics of the differences between the various signals used in the test. In conversations following the tests, many participants noted a difficulty in discriminating between the “A” and “B” signals for the speech sound file. This corresponds with the fact that the two lowest scores for the stimuli were for the two examples of speech with reverberation. Generally, comments indicated that the xylophone signal differences were most evident due either to a presence or lack of slap-back echo (corresponding to the specular reflections) or a timbre change (due to increased high frequency components in the diffused model). Comments regarding the cello signals indicated that resonances in the low frequencies (a result of the comb filtering specular reflections) proved to be the strongest indicator. Two subjects, however, noted an change in the apparent distance to the instrument.

The primary conclusion of this test is that subjects are easily able to distinguish the difference between audio signals processed using the two models, whether in the presence of a reverberant tail or not. This ensures a higher degree of reliability of the data obtained from the second test.

5.3 Test 2: Mix preference test

The primary purpose of the listening tests is to determine whether subjects prefer the diffuse reflection model over the specular equivalent, or some mix of the two. This is achieved through a blind test in which subjects are able to select a relative balance between a fully specular and fully diffuse reflection model in real time.

Instead of a continuously variable balance between the two reflection models, the mix was quantized into seven possible responses corresponding to the levels listed in Table 5.5. The gain values used for these mixes were calculated using a constant power function using cosine and sine functions with 15° increments. This ensured that there was a constant power at the listening position, thus eliminating level differences as a contributing factor. This was verified with the sound pressure level meter described in Section 5.1.3 using a pink noise sound source. After the use of an A-weighting filter, there was less than a 0.1 dB difference between the various mix values listed in Table 5.5.

5.3 Test 2: Mix preference test

| Mix | Power Ratio
(Diff / Spec) | Diffuse gain | Specular gain | Total (dB _{spl} A) |
|-----|------------------------------|--------------|---------------|-----------------------------|
| a | 0% / 100% | 0.000 | 1.000 | 58.7 |
| b | 17% / 83% | 0.259 | 0.966 | 58.7 |
| c | 33% / 67% | 0.500 | 0.866 | 58.7 |
| d | 50% / 50% | 0.707 | 0.707 | 58.7 |
| e | 67% / 33% | 0.866 | 0.500 | 58.7 |
| f | 83% / 17% | 0.966 | 0.259 | 58.7 |
| g | 100% / 0% | 1.000 | 0.000 | 58.7 |

Table 5.5 List of specular / reflection gains for the seven mixes used in the second listening test. The far right column shows the measured total sound pressure level output of the five channels using the pink noise signal described in Section 5.1.3 without reverberation.

Figure 5.3 shows a screen shot of the display used for the second listening test. In it, subjects were asked to use the left and right arrow buttons or the corresponding cursor keys on the keyboard to alter the signal to their desired mix. Subjects were given no prior indication of the audible differences between the two signals, however, all were told that there were two different signals that were identical to the “A” and “B” signals from the first test in the previous week. In order to avoid any visual cues, the balance was adjusted using the left and right cursor keys on the computer keyboard without feedback on the computer monitor. In addition, no cue was included to indicate that a mix of completely specular or diffuse signals had been reached. It should also be noted that, for each stimulus, the initial balance was randomly chosen from the seven possible mixes.

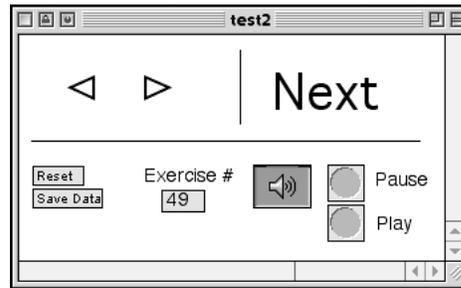


Figure 5.3 Screen shot of test window used in mix preference test

5.3.1 Results

The results of the second listening test are listed in Table 5.6. This table lists a number of statistics regarding the results of the test which should be explained. The responses from the listening test were tabulated and converted from a pressure amplitude gain to a relative power level in order to list a mix “percentage.” The values listed in the “Mean Power” column are the averages of the responses converted to a power scale and listed as the level of the diffuse component. The “skewness coefficient” is an indicator of the asymmetry of the distribution of the responses. Negative values indicate that a longer “tail” in the distribution curve is on the left of the mean. Positive numbers indicate it is on the right, and 0 indicates that the distribution is symmetrical around the mean.

5.3.1 Results

| Stimulus | Sound file | Reverb |
|----------|------------|--------|
| 1 | Speech | No |
| 2 | Speech | Yes |
| 3 | Xylophone | No |
| 4 | Xylophone | Yes |
| 5 | Cello | No |
| 6 | Cello | Yes |

Table 5.6 List of the stimuli numbers for reference in Table 5.7 and Figure 5.4

| Stimulus | Mean Power (Diffuse) | Standard Deviation | 99% confidence interval | Skewness coefficient |
|----------|----------------------|--------------------|-------------------------|----------------------|
| 1 | 0.73 | 0.33 | ± 0.09 | -1.15 |
| 2 | 0.52 | 0.37 | ± 0.10 | 0.02 |
| 3 | 0.79 | 0.32 | ± 0.09 | -1.58 |
| 4 | 0.69 | 0.34 | ± 0.09 | -0.88 |
| 5 | 0.46 | 0.33 | ± 0.09 | 0.05 |
| 6 | 0.49 | 0.35 | ± 0.10 | 0.10 |

Table 5.7 Statistics of the responses from the listening test. Note that all values are based on the level of the diffuse component converted into a mix percentage (power level from pressure gain).

A number of conclusions can be drawn from these data. Firstly, note that the speech without reverberation (subsequently labelled “dry speech”) and both xylophone stimuli correspond to higher preferred levels than the speech with reverberation (or “wet speech”) and both cello stimuli. This, in and of itself is not indicative of a general trend since the distribution of the responses, indicating the degree of agreement between the subjects,

must be taken into account as well. Using the means and the interquartile ranges displayed in Figure 5.6, it can be seen in the case of speech with reverberation and cello, that the stimuli have wider distributions and thus indicate that there was less agreement for a single value for the mix. In comparison, both xylophone stimuli have smaller interquartile ranges, thus indicating that the level of agreement between subjects is higher.

Two conclusions can be drawn from this distribution. The first is that, for the xylophone and dry speech stimuli, and thus extrapolating to any transient program material, a predominantly diffused model is preferred over a more specular model. This corresponds with informal comments from many subjects following the test regarding the unpleasant “slapback” echo effect heard in the specular model which is not audible in the case of steady-state signals. The second conclusion is that, although there is less agreement on a single mix position for the cello and wet speech samples, this is partially an indicator of personal preference and simple noise in the data. One rewarding aspect of this analysis is that there was no firm agreement against the diffused model and that, in some cases (such as cello with reverberation) the preference is a mix of both specular and diffused reflections.

5.3.1 Results

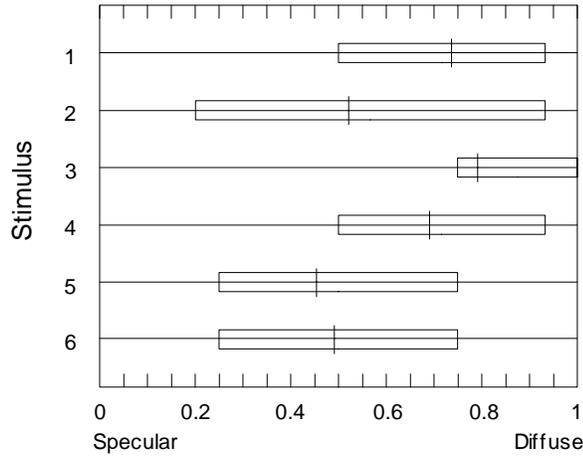


Figure 5.6 Box and whisker plots showing the means and interquartile ranges for the six stimuli.

The relatively high degree of negative skewness in the two xylophone stimuli, evident in both the table listings and the box and whisker plots, point to the high number of responses indicating a preferred 100% diffused model mix. The level of this response was reduced by a number of much lower-level responses resulting in a lower mean.

| Stimulus | Subject "a" | | Subject "b" | |
|----------|----------------------|--------------------|----------------------|--------------------|
| | Mean Power (Diffuse) | Standard Deviation | Mean Power (Diffuse) | Standard Deviation |
| 1 | 0.90 | 0.09 | 0.90 | 0.11 |
| 2 | 0.92 | 0.09 | 0.16 | 0.25 |
| 3 | 0.54 | 0.29 | 0.96 | 0.07 |
| 4 | 0.54 | 0.31 | 0.98 | 0.03 |
| 5 | 0.58 | 0.14 | 0.78 | 0.21 |
| 6 | 0.50 | 0.29 | 0.82 | 0.29 |

Table 5.8 Statistics of the responses from the two subjects with the highest degrees of reliability.

5.3.1 Results

Of note are the discrepancies between individual preferences. Table 5.8 shows the data collected from the two subjects with scores indicating the highest degrees of reliability of the group. Note that, although the standard deviation values for both subjects are lower than those of the entire group, they do not necessarily correspond to the same conclusions. As can be seen in the table, subject “a” has the highest degree of reliability for stimuli using the speech signal, whereas, like the total group, subject “b” has the highest degree of reliability for the xylophone. Another characteristic to point out is the difference in opinion of the two subjects regarding the preferred mix for the xylophone signals. Note that, while subject “a” preferred a nearly equal balance of specular and diffuse reflections, subject “b” preferred a nearly completely diffused model. Finally, one interesting characteristic is the difference in subject “b’s” response for the speech stimuli. There is a remarkable difference in preference in the speech stimuli with and without the reverberation, possibly indicating a difference in intelligibility which should be evaluated using the MLSSA measurement system following the future implementation of the system as a real-time processor.

5.4 Analysis

A number of conclusions can be drawn from the results of the listening test. The first is that the two models are distinguishable by experienced listeners, with a varying degree of difficulty according to the program material chosen. The second is that, while there is dispute regarding the aesthetic merit of the diffused model for all sound sources, there is an indication that the inclusion of a balance control over the relative mixes of the specular and diffused models of early reflections should be implemented on artificial reverberation systems. Additionally, there is no question, in the case of program material with a large transient component, such as percussion instruments, that the diffused model is preferable to the specular model by most listeners.

One particular issue of note is the low order of reflection that was used in the listening examples. As will be discussed in the following chapter, the use of isolated first order reflections is inadequate for any usage, consequently, although the model shows promise as a new method of simulating early reflections, further development is needed to extend the algorithm to higher reflection orders.

6 Conclusion

Although listening tests indicate that the procedure for synthesizing early reflections outlined in this dissertation are either an improvement over or should be used in conjunction with existing models, there are a number of possible avenues for development and improvement to be pursued in the future. These developments include both extensions to the model as well as refinements in various components and aspects regarding to the implementation.

6.1 Future developments

While it has been proven using the listening tests that a mix of specular and diffused reflection models results in preferable reflection characteristics than a typical perfectly specular reflection model, there are a number of improvements that would increase the quality of the system, both on an aesthetic and an ergonomic level.

6.1.1 Aesthetic improvements

The primary limitation of the system is stated in the title of this dissertation: the model does not include the third dimension of height. Preliminary investigations performed in the MARLab using image models of rooms with perfectly specular reflective surfaces indicate that the inclusion of a height component in a synthetic room model greatly improves the beauty and realism of the resulting sound, even when reproduced using a two-dimensional loudspeaker configuration.

The second principal limitation of the system is the fact that the model has been developed exclusively for the first order reflections. Calculating a second-order reflection from two diffusive surfaces dramatically increases the combinatorial complexity of the system. This is because a diffusive surface acts effectively as multiple sound sources simultaneously. Consequently, the number of discrete sound source locations generated by a first reflection which would be required to compute any higher order reflection using the system described in Chapter 3 would be prohibitive.

There are a number of less significant constraints in the system including:

6.1.1 Aesthetic improvements

1. The absence of surface models other than a Schroeder diffuser. This is primarily a limitation of the implementation rather than the algorithm, since the only modification to the system necessary to correct this constraint would be the replacement of the impedance calculation.
2. The system should be evaluated against a real model of a Schroeder diffuser in order to determine the validity of the physical model. This would subsequently determine the legitimacy of the system for predictive acoustics systems.
3. Eventually, an ultimate goal would be to develop an algorithm that creates the entire reverberation tail in the same manner as this general method – a macroscopic result from an assembly of microscopic models.

One possible long-term method of aesthetic evaluation would be to use a system similar to that used by corporations in modern product development. If the algorithm were to be implemented as a real-time processor, it could be installed as a “beta-testing” component for use in a recording studio for direct comparison with commercially-available digital reverberation devices. In this situation, feedback from users could easily point to attributes of the algorithm that require additional refinement.

6.1.2 Ergonomic improvements

There is one principal improvement required for the entire system: the implementation of the algorithm in a real-time system. While such development work lies outside the scope of this dissertation, it is the necessary next step in the progress of the system. The creation

of the sound samples used in the listening tests required hours of computation time (including calculation of the impulse responses and convolution) on a 350 MHz Apple Macintosh G3 computer. This time is prohibitive to the refinement of the model due to the long wait times between manipulation of the parameters and the resulting audio output.

Figure 6.1 shows a proposed method of implementing the system as a real time processor. The impulse response of each well in the Schroeder diffuser is simulated by a physical model in the form of a recursive comb filter, ensuring that the characteristics of the filter in the feedback loop result in a resonance pattern matching the well itself. Since the Schroeder diffuser is periodic with a number of wells determined by the value “N” this determines the required number of well models. Further efficiency can be obtained by simulating all wells with similar dimensions within the period with the same physical

model, resulting in a total of $\frac{N+1}{2}$ well simulators. The signal from the sound source impinging on the mouth of each well is simulated using a delay with a gain which is the product of the gains resulting from the instrument and microphone directivities as well as the gain due to propagation to and from the reflective surface. The frequency-dependent characteristics of the instrument and well radiation patterns, as well as the polar sensitivity pattern of the microphone are incorporated into the filter placed in the signal path between the output of the delay and the input of the well simulator. Since the wave from the sound source reflects off a number of wells of the same size, each well simulator

would be fed the summed output of a multitap delay. The number of required taps from this delay would be equal to the number of similarly-sized wells on the reflective surface.

One possible problem with this model is the lack of density in comparison with the model described in Chapter 3. Whereas, in the present system, a local well impulse response is determined to begin at each sample in the impulse response, the real-time implementation would have only one representation of each well. Consequently, the real-time implementation would have to be evaluated against the non-real time prototype to ensure that a minimum drop in sound quality would be incurred.

6.1.2 Ergonomic improvements

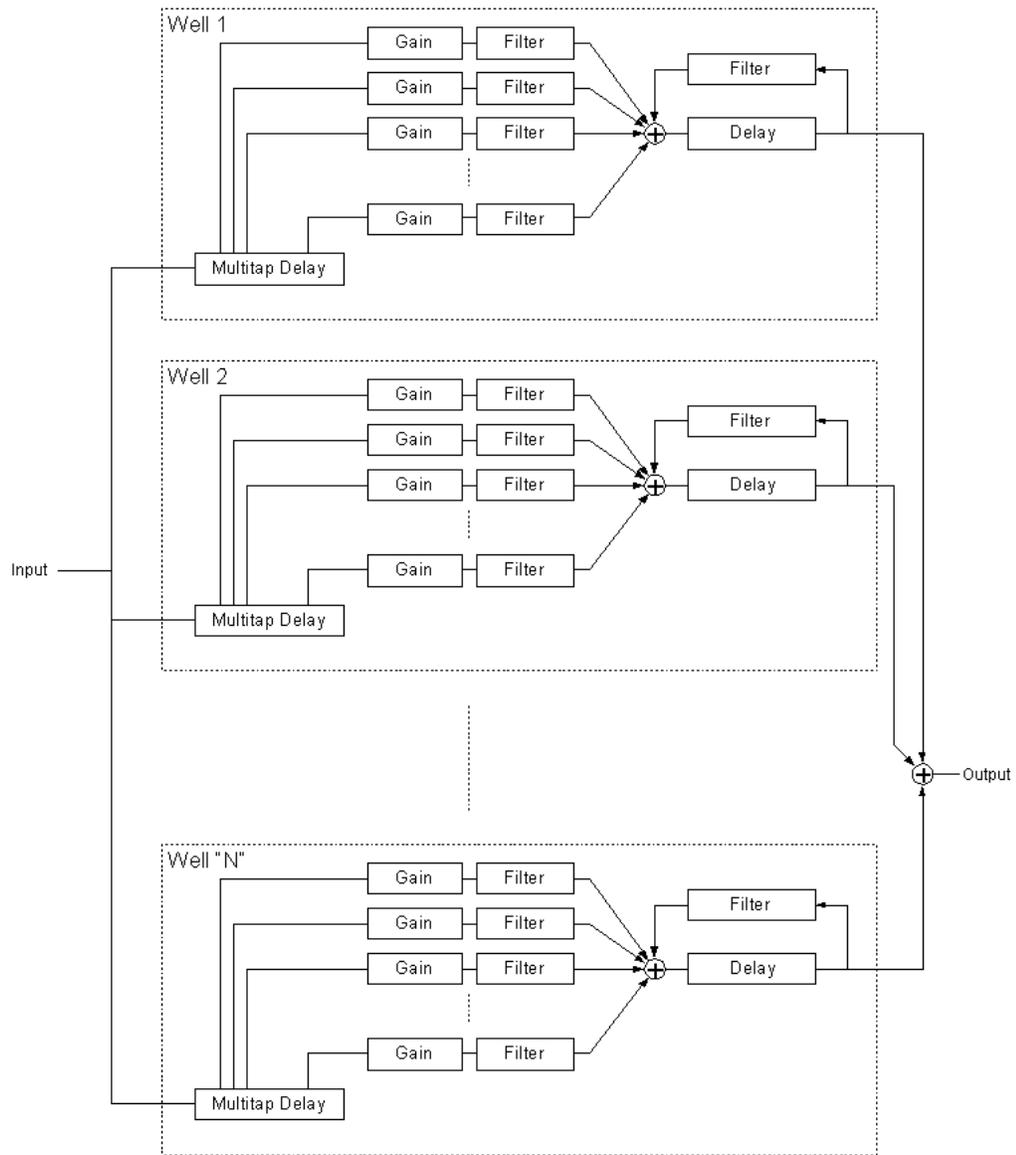


Figure 6.1 Proposed algorithm for real-time implementation using a recursive delay to simulate the individual diffuser wells.

A considerably less significant issue to be addressed in further development work for the ergonomic response of the system includes refinement of the graphical user interface. This aspect was not developed for the present implementation since the work concentrated on the development of the algorithm rather than the platform itself.

6.1.3 Technical considerations

There are a number of technical issues which require consideration and further improvement in later versions and implementations of the algorithm.

The first issue is a question of the accuracy of the physical model used to produce the impulse responses off the individual wells in the Schroeder diffuser. In particular, there are two points which require refinement over the implemented model, the angle of incidence of the pressure wave impinging on the entrance of the well and the frequency-dependent directivity of the re-radiated power.

The model used a simplified model to calculate the acoustic impedance at the entrance to the well. This model assumes that the incoming wave is traveling along the normal to the well mouth. While this is indeed the case at one location on the reflective surface, it is not the case for all other points of reflection. As a result, one recommended refinement of the model would be to include an appropriately calculated impedance for the reflective surface according to the angle of incidence of the pressure wave.

Similarly, the model does not include the issue of the directivity pattern of the radiating well. The mouth of the well is assumed in the model to be a circular piston which would

have a radiation pattern which was frequency-dependent according to the diameter of the well. The immediate result of the inclusion of this directivity function would be the loss of high-frequency content in the reflected signal, reducing the relative differences between the low and high frequency bands in the frequency response graphs of the diffuse reflection shown in Chapter 4. Equation 6.1 shows this function which demonstrates that the higher the frequency, the greater the directivity.

$$\frac{p_{\vartheta_r}}{p_0} = 1 - \frac{2\pi}{8c} f r \sin \vartheta_r \quad (6.1)$$

where p_{ϑ_r} is the pressure of the radiated wave at the angle of reflection ϑ_r , p_0 is the pressure of the same wave along the normal at the same distance of measurement and r is the radius of the well (Jordan 1963).

An issue of note in the implementation of the system is the size of the IFFT used to convert the impedance response into an impulse response. For the sake of speed in this implementation, an IFFT length of 8192 points was used. The result of this rather small length was the creation of artifacts at the end of each well's individual impulse response. These artifacts typically took the form of a small increase in the level of the resonant frequency of the well at the end of the impulse response. In the tested implementation, the increase in level of the noise burst was calculated to be approximately 40 dB below the peak level of the well's initial impulse response level at a frequency close to the Nyquist. The use of a larger IFFT length reduces the levels of these artifacts but also requires a corresponding increase in calculation time.

Additional minor possible developments for the system include the implementation of a binaural output for auralization, user-selectable wall materials, specific locations for the diffusion panels on a reflective surface (rather than entire surfaces being comprised of the diffuser wells) and the inclusion of high frequency loss due to air absorption for large spaces.

One final technical issue that remains to be evaluated is the question of perception of space. The listening tests described in Chapter 5 were used to prove that the algorithm is an improvement on existing synthetic early reflection models on a simple aesthetic level, however, it remains to be seen whether listeners derive a more accurate impression of the synthetic space. There are two possible avenues for conducting this investigation. The first is to run a series of listening tests in which subjects are first evaluated on their ability to determine room size based on acoustic characteristics, then, using the high scorers from this group, evaluate the impression of room size in the specular and diffuse models. The second investigation would consist of a direct comparison between recordings produced in synthetic and real spaces constructed of specular and diffusive surfaces.

6.2 Conclusion

While the model has been proven to be an improvement over existing methods of generating specular early reflections, much work remains to refine the model to create a system that is both aesthetically and ergonomically acceptable while maintaining a feasible level of computational requirements. As processing power inevitably increases, the challenge will remain to improve the system to provide a usable tool for recording engineers, sound designers and composers, however, the foundation inarguably exists to build a new model of synthetic reverberation. Contrary to Dostoevsky's observation, it may indeed be possible to use the analysis of the parts to improve both the accuracy and the beauty of the whole.

Appendix A: Effects processors parameters

This section lists the parameter settings for the two reverberation units measured and analyzed in Section 1.7.2.

A.1 t.c. electronic M3000 parameters

The measurements of the t.c. electronic M3000 were accomplished by creating an .aiff file consisting of a series of 0's and a single sample of value 1. This Dirac impulse was sent from Max/MSP through a Mark of the Unicorn 2408 to the reverberation unit using an S/PDIF digital audio connection. The digital input of the device was locked to the 2408 with a sampling rate of 44100 Hz. The output of the device was returned to the 2408 through an S/PDIF connection and this incoming signal was recorded to an .aiff file using Max/MSP. This file was converted to .wav format using the SoundHack software package. All subsequent analysis of the impulse response was done using the .wav file imported in MATLAB.

The parameter configuration of the M3000 is listed below in Table A.1.

A.1 t.c. electronic M3000 parameters

| | | | |
|----------------------|-------------|--------------------------|----------|
| I/O | | Early Reflections | |
| Input | SPDIF | EarlyType | ConcHall |
| Channel | L | EarlySize | Large |
| Mix | Mix | EarlyBal | Centre |
| Routing | | HiColor | default |
| Dual Mono (Engine 1) | | LoCut | 20 Hz |
| Engine 1 | | Reverb | |
| Patch | Nice Hall 1 | RevType | Smooth |
| Decay | 2.0 Sec | Diffuse | Default |
| EarlyLev | 0 dB | RevBal | Center |
| RevLev | 0 dB | HiCut | 20.0 kHz |
| Mix | 100 % | HiSoften | default |
| OutLevel | 0 dB | HiDecay | 1.00 |
| RevDelay | 200 ms | HiXover | 1.30 kHz |
| PreDelay | 0 ms | MidDecay | 1.00 |
| | | MidXover | 298.5 Hz |
| | | Lmid Decay | 1.00 |
| | | Lo Xover | 91.73 Hz |
| | | Low Decay | 1.00 |
| | | Low Damp Freq | 183.0 Hz |
| | | Lo Damp | 0.0 dB |
| | | Reverb Modulation | off |
| | | Space Modulation | off |

Table A.1 Parameter settings for M3000 measurements

A.2 Lexicon 480L parameters

Since the 480L in McGill University’s Studio A is not equipped with the digital I/O option, but does have balanced analog inputs, the system was measured using the MLS test in the Audio Precision System Two. The two balanced analog inputs and outputs were connected directly to the XLR inputs and outputs of the 480.

The parameter configuration of the 480L are listed below in Table A.2. Note that many of these parameters are not applicable to the measurement.

| | | | |
|--------------------|---------------|-------------------------------------|--------|
| Page 1 | | Page 3 – Preecho Levels | |
| RTM | 0.02 sec | L → L | Full |
| Shape | 0 | R → R | Full |
| Spread | 3 | All others | Off |
| Size | 4 m | | |
| HF Cutoff | Full range | Page 4 – Preecho Delay times | |
| Predelay | 300 msec | L → L | 8 msec |
| | | R → R | 8 msec |
| | | All others | 0 msec |
| Page 2 | | | |
| Bass Multiply | x 1.00 | | |
| Crossover | 752 Hz | | |
| RT HF Cut | Full range | | |
| Diffusion | <i>varied</i> | | |
| Decay Optimization | Reverb 7 | | |
| Mix | All effects | | |

Table A.2 Parameter settings for 480L measurements

Appendix B: Technical details of sound source recordings

The sound sources used for all IACC measurements and listening tests are excerpts from a compact disc produced for the Archimedes project by Bang & Olufsen. Although the disc contains both anechoic and semi-reverberant samples, only anechoic sources were used for this research.

| | |
|---------------------------|--|
| Compact Disc Title | Music for Archimedes
CD B&O 101 |
| Copyright | Bang & Olufsen 1992 |
| Female Speech | Danish
Track 8 |
| Source Height | 1.50 m |
| Microphone Height | 1.49 m |
| Distance to microphone | 0.5 m |
| Microphone | Brüel & Kjær 4003
Pointing at sound source
Speaker turned 30° away from microphone |
| Cello | Variation and Theme No. 2, Weber
Track 22 |
| Source Height | 0.80 m |
| Microphone Height | 1.40 m |
| Distance to microphone | 1.1 m |
| Microphone | Sennheiser MKH40 P48
pointing at upper half of cello |

| | |
|------------------------|-------------------------------------|
| Xylophone | Sabre Dance, Khachaturian |
| | Track 27 |
| Source Height | 1.40 m |
| Microphone Height | 2.45 m |
| Distance to microphone | 1.20 m |
| Microphone | Sennheiser MKH40 P48 |
| | Pointing to rear edge of instrument |

Appendix C: MARLab acoustical characteristics

McGill University's Multichannel Audio Research Laboratory or MARLab is situated in the basement of Redpath Hall. The room is designed for research into the characteristics of multichannel reproduction systems and, as such, is ideally suited for the described evaluation tests. The room has a total area of 35.51 m^2 and a height of 2.22 m resulting in a total volume of 78.83 m^3 . A heavy velour curtain is hung approximately 10 cm from the walls and along the front of the alcove as is seen in Figure C.1. This results in a total enclosed space of approximately 72.88 m^3 . The listening area is surrounded by an acoustically transparent but visually opaque white curtain preventing any visual cues such as loudspeaker location. The ceiling is comprised of standard acoustical tiles hung in a T-bar support. Two layers of 5 cm thick mineral wool rest on the upper side of these panels to reduce low frequencies and resonances between the tiles and the plaster upper ceiling at a height of 3.00 m . In addition, tuned membrane absorbers are attached to the South, South-West and North-West walls to reduce characteristic room resonances.

The room is supplied by a dedicated air conditioning system with lined ductwork to reduce airflow noise as much as possible. In addition, the room has a small machine area to isolate intrusions such as computer fan and hard drive noise.

Typical noise floor measurements for the room are below $27 \text{ dB}_{\text{spl}}$, A-weighted, predominantly a result of mechanical noise from a power transformer directly above the laboratory on the second floor of the building and external noise such as traffic. (Note

that, in a recent power outage, the noise floor without the noise of the transformer or air conditioning system was measured to be 20.5 dB_{spl} A-weighted.) Reverberation times were measured with the results detailed below.

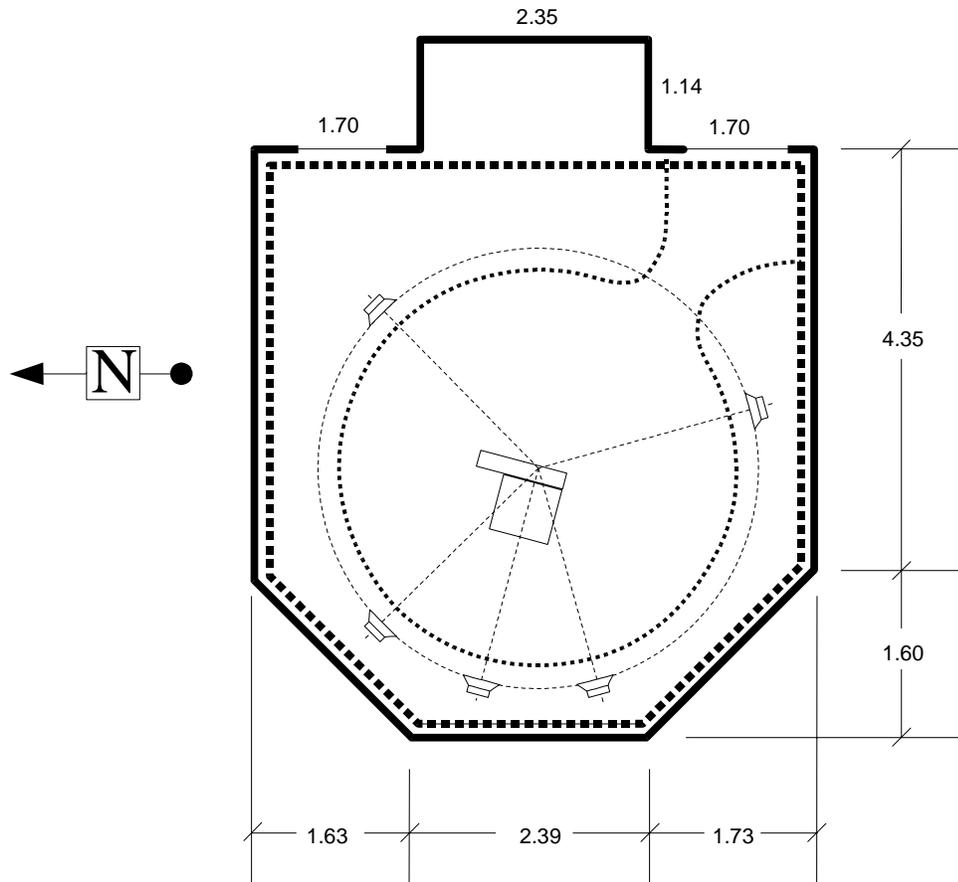


Figure C.1 Floorplan of MARLab. The heavy dotted line parallel to the walls denotes the velvet curtain. The lighter dotted line is an acoustically transparent but visually opaque white curtain preventing subjects from seeing the loudspeaker locations. All measurements are in metres.

The measurement of the reverberation times of the MARLab were performed using a DRA Laboratories Maximum Length Systems Sequence Analyzer or MLSSA. The analog

output of the MLSSA was connected directly to the RCA input of a Bang & Olufsen Beolab 4000 two-way self-powered loudspeaker (Type number 6638 R, Serial number 12614491). The loudspeaker was placed in the North-East corner of the room, resting on an 11.5 cm high plywood cable trough at the cable entrance to the machine room. The microphone was placed near the ceiling panels in the opposite corner of the room, at the junction of the South and South-West walls. The microphone used for this measurement was a Brüel & Kjær 4006 P48 microphone (Serial number 1533840) connected to a Tascam DA-P1 portable DAT machine (Serial number 70321T) whose output was connected directly to the input of the MLSSA.

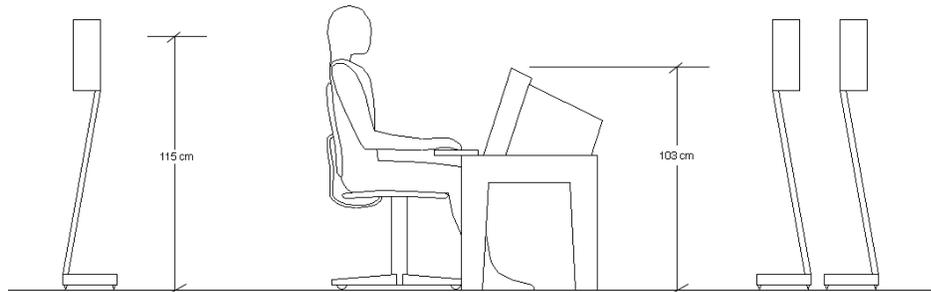


Figure C.2 Side view of loudspeaker, computer monitor and subject. Figure drawn to scale for 183 cm tall male.

The MLSSA Acquisition parameters were set to the values shown in Table C.1:

| Parameter | Value |
|---------------------|----------------|
| Stimulus Level | 0.1 V |
| Pre-Averages | 16 |
| Bandwidth | 10 kHz |
| Antialiasing Filter | Butterworth |
| Length | 32,768 Samples |

Table C.1 MLSSA parameter settings for room reverberation time measurements.

Table C.2 shows a partial list of the results of the Calculate Acoustics command in the MLSSA. These data show the results for seven octave bands ranging from 125 Hz to 8 kHz. The Signal to Noise or S/N values are listed in dB to indicate the reliability of the data. The MLSSA manual recommends that all parameters in frequency bands with signal to noise ratios of less than 10 dB be “regarded as suspect” (Rife 2000) however, this is not an issue as the minimum signal to noise ratio in these measurements is 32.3 dB in the 8 kHz band. Note that all calculations are made automatically by the system, compensating for noise levels which are measured from the portion of the recorded impulse response which occurs before the arrival of the initial impulse.

The Early Decay Time or “EDT” is the decay time of the reverberation for the first 10 dB of decay. The RT-20 values are extrapolated RT₆₀ values (the time it takes the reverberation to decay 60 dB) calculated using the 20 dB “window” of -5 dB to -25 dB. In theory, the decay curve of the reverberation in the -5 dB to -25 dB window should exactly

fit a perfectly straight line. In practice, however, there is some deviation in the decay curve. This deviation is measured as a correlation coefficient which is given in the last line of the table. A value of -1.000 indicates that the decay curve matches the theoretical curve exactly. Larger numbers indicate less reliable data with values greater than -0.950 indicating a nonlinear reverberant decay and therefore suspect data. Again, the results of this measurement are all valid with a worst-case correlation coefficient of -0.967 in the 250 Hz octave band.

| | 125 Hz | 250 Hz | 500 Hz | 1 kHz | 2 kHz | 4 kHz | 8 kHz |
|-------------------|---------------|---------------|---------------|--------------|--------------|--------------|--------------|
| S/N (dB) | 59.8 | 44.7 | 33.1 | 37.1 | 33.1 | 32.6 | 32.3 |
| EDT (ms) | 242 | 209 | 168 | 98 | 65 | 66 | 75 |
| RT-20 (ms) | 270 | 172 | 177 | 100 | 110 | 111 | 88 |
| -5, -25 | -0.976 | -0.967 | -0.989 | -0.997 | -0.995 | -0.994 | -0.995 |

Table C.2 Partial results of MLSSA **Calculate Acoustics** command

Appendix D: Monitoring system response characteristics

The calibration of the reproduction system involved two stages. In the first, the frequency response of each loudspeaker was individually measured in a pseudo-anechoic environment using a DRA Laboratories MLSSA. For each measurement, the loudspeaker was positioned on a pedestal located in the centre of Redpath Hall such that the tweeter was 2.56 m off the floor, the closest reflective surface. A Brüel & Kjær 4006 Microphone (Serial number 1533840) with a silver grid was positioned 1 m in front of the loudspeaker at the same height as the tweeter with the loudspeaker on axis to the microphone. Figure D.1 shows a photograph of the arrangement. The output of the microphone was pre-amplified using a Mackie 1604-VLZ analog mixer whose RCA tape output was fed to the MLSSA input. The output of the MLSSA was connected directly to the RCA input of the loudspeaker.

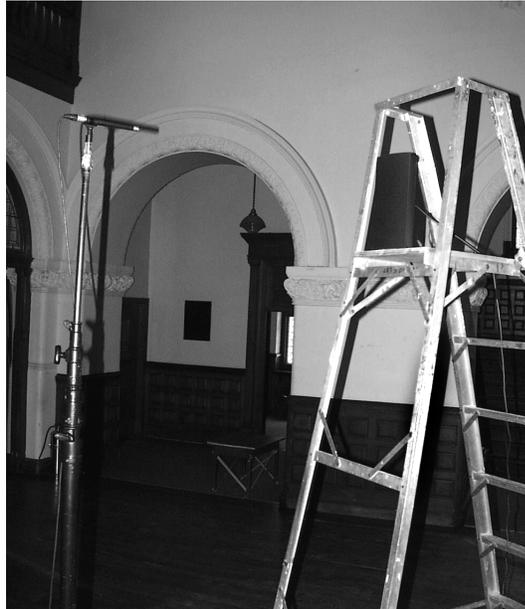


Figure D.1 Photograph of pseudoanechoic loudspeaker measurement configuration in Redpath Hall

In order to make an accurate frequency response measurement of an electroacoustic device, the measurement must take place in an anechoic environment to avoid the influence of the surrounding acoustic environment. One principal advantage of the MLSSA system is that it is able to selectively window the impulse response measurement in the temporal domain, thus excluding any reflections from surrounding sources and resulting in an effectively anechoic measurement. Figure D.2 shows a portion of the measured impulse response of the centre loudspeaker. Note that the first early reflection is visible at approximately 15 ms. Slightly to the left of this reflection is the cursor indicating the maximum limit of the time window for all subsequent frequency measurements.

The bandwidth of the MLSSA system for these measurements was 20 kHz using a Butterworth antialiasing filter. Since the recommended sampling rate with this filter is four times the bandwidth, the resulting sampling rate was 80 kHz. Each measurement is the result of 16 concurrent pre-averages in order to increase the signal to noise ratio of the measurement system to its maximum possible value. Table D.1 displays the parameter settings for all pseudo-anechoic loudspeaker measurements.

| Acquisition | | Units | |
|----------------------------|-------------------------|-----------------|--------------------|
| Mode | Cross-correlation | Acquisition | 1 Volt / Volt |
| Length | 8192 samples (102.4 ms) | Stimulus | 1 Volt / Volt |
| Sample rate | 80 kHz (12.5 μ s) | Stimulus | |
| Concurrent pre-average | 16 | Burst MLS | |
| Auto range | Enabled | Amplitude | \pm 0.2051 Volts |
| Antialiasing filter | | Rep rate | 4.883 Hz |
| Butterworth | | Period | 16863 Samples |
| Bandwidth | 20 kHz | | 204.8 ms |
| Gain | 1 (\pm 5 V range) | Trigger | |
| Type | Stimulus Trigger | | |
| Delay | 0 samples (0 ms) | | |

Table D.1 Parameter settings of the MLSSA software for all pseudoanechoic measurements

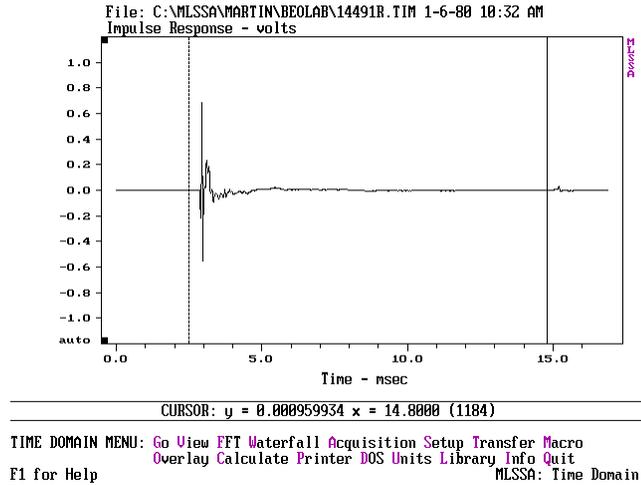


Figure D.2 Centre loudspeaker time measurement showing first reflection from floor outside of cursor. Note that the date of measurement is incorrect – the result of a simple fix to avoid Y2K errors.

This temporal measurement was used to calculate a frequency domain response using the MLSSA **FFT Execute** function. The FFT size was 8192 points, using a half Hanning windowing function with 0.333-octave smoothing with the marker and cursor position as indicated in Figure D.2. The resulting frequency responses of the five loudspeakers are displayed in Figures D.3 to D.7 inclusive. Note that these frequency response plots do not necessarily indicate the response of the system at the listening position in the monitoring room. They are used only to determine the relative matching of the loudspeakers in an anechoic environment. Note as well that, since the total time of the window used to calculate each of these frequency-domain responses is approximately 12 ms, the displayed responses below approximately 83 Hz are not reliable.

Appendix D: Monitoring system response characteristics

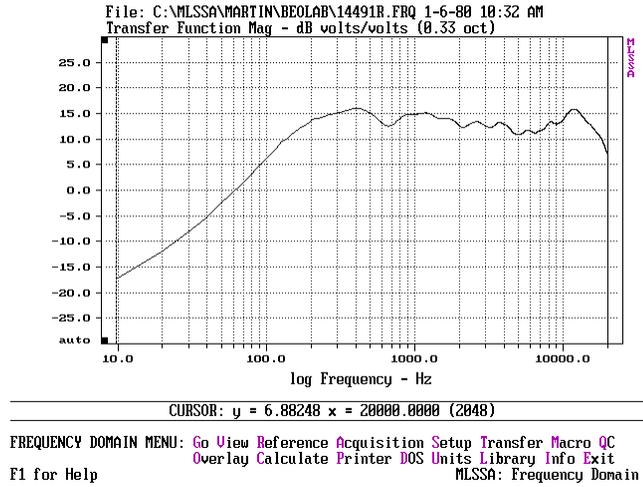


Figure D.3 Pseudo-anechoic one-third octave smoothed frequency response of centre loudspeaker

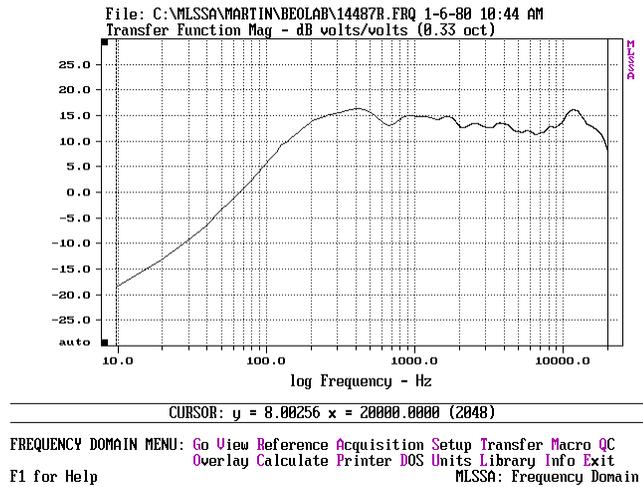


Figure D.4 Pseudo-anechoic one-third octave smoothed frequency response of left loudspeaker

Appendix D: Monitoring system response characteristics

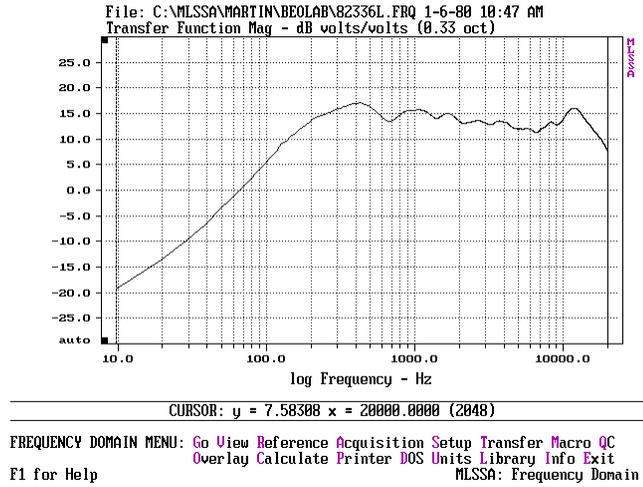


Figure D.5 Pseudo-anechoic one-third octave smoothed frequency response of right loudspeaker

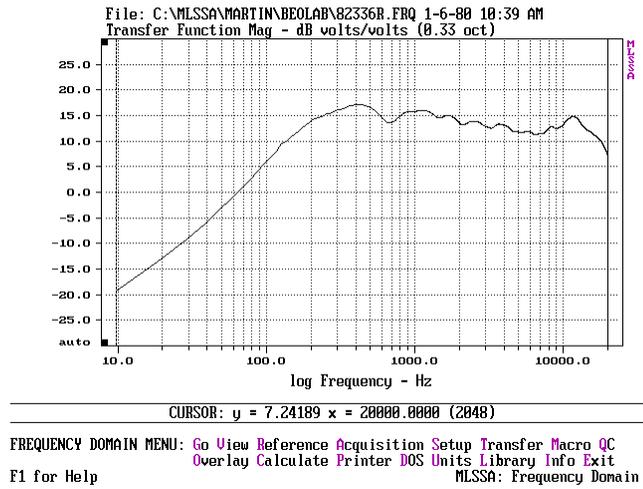


Figure D.6 Pseudo-anechoic one-third octave smoothed frequency response of left surround loudspeaker

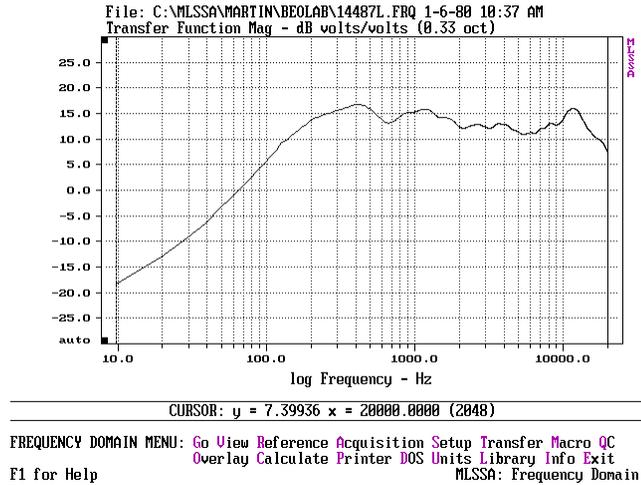


Figure D.7 Pseudo-anechoic one-third octave smoothed frequency response of right surround loudspeaker

As can be seen in the five included plots, the five loudspeakers are extremely well-matched with similar response characteristic curves. There are some minor deviations in the level of the peak at approximately 12 kHz with a maximum difference between any two loudspeakers of 1.5 dB.

In order to calibrate the system in the listening room, the entire system must be measured as a single device. This procedure consisted of initially measuring the relative responses of the five electrical paths from a single input to the MotU 2408, routed to the five output channels using Max/MSP back through the analog outputs of the 2408 to the Yamaha 03D. Using MLSSA, the relative responses of these five signal paths were measured with a bandwidth of 20 kHz. All channels were found to have the same characteristic frequency response shape (a gentle rise of +0.5 dB in the high frequencies and a low frequency rolloff of approximately 1 dB per octave from 10 Hz to 50 Hz) and were

matched in amplitude within 0.5 dB. This difference in level was compensated in a later procedure.

The total signal path, encompassing the above system and including the individual loudspeaker outputs and room response was measured by performing the same measurement with the input to the MLSSA from a microphone located at the listening position. A Brüel & Kjær 4006 P48 microphone (Serial number 1533840) connected to a Tascam DA-P1 portable DAT machine (Serial number 70321T) was used for signal acquisition. The microphone was aimed vertically to avoid off-axis response differences between the various loudspeakers. The height of the microphone diaphragm was 108 cm from the floor.

This configuration was used for two measurements. The first was to investigate the individual levels of the initial impulse in order to calibrate the relative levels of the five output channels. This was done using the parameters listed in Table D.2. Each output channel was measured and the corresponding output fader on the Yamaha 03D was adjusted to achieve the closest possible match in level to the centre loudspeaker. These calibration values on the output busses were maintained as static adjustments throughout all listening tests.

| Acquisition | | Units | |
|----------------------------|-------------------------|-----------------|--------------------|
| Mode | Cross-correlation | Acquisition | 1 Volt / Volt |
| Length | 4096 samples (204.8 ms) | Stimulus | 1 Volt / Volt |
| Sample rate | 20 kHz (50 μ s) | Stimulus | |
| Concurrent pre-average | 4 | Burst MLS | |
| Auto range | Enabled | Amplitude | \pm 0.4922 Volts |
| Antialiasing filter | | Rep rate | 1.221 Hz |
| Butterworth | | Period | 16863 Samples |
| Bandwidth | 5 kHz | | 819.2 ms |
| Gain | 2 (\pm 2.5 V range) | | |
| Trigger | | | |
| Type | Stimulus Trigger | | |
| Delay | 0 samples (0 ms) | | |

Table D.2 Parameter settings of the MLSSA software for the lab system level calibration measurements

Table D.3 shows the results of these measurements after calibration. Note the extremely matched levels with a maximum level difference of 0.31 dB and 50 μ sec between any two channels.

| Channel | Arrival Time (ms) | Amplitude (V) | dB _{centre} |
|----------------|-------------------|---------------|----------------------|
| Left | 16.500 | 0.318677 | 0.04 |
| Centre | 16.500 | 0.317379 | 0.00 |
| Right | 16.550 | 0.322630 | 0.14 |
| Left Surround | 16.500 | 0.312824 | -0.13 |
| Right Surround | 16.550 | 0.311239 | -0.17 |

Table D.3 Relative levels of initial impulse at listening position in MARLab. Note that the right column is a calculation of the relative level of the channel referenced to the centre loudspeaker.

The second measurement was an investigation of the relative frequency responses of the five channels in the system, including room response. No attempt was made to use this as a calibration method as no adjustments to the system were applied as a result of this measurement. The test was performed for documentation purposes only.

The FFT calculations used an FFT size of 8192 points, with a total time window of approximately 178 ms beginning at 0 ms. A half-Hanning window was used with 0.333 octave smoothing. All other parameters are listed in Table D.4.

| Acquisition | | Units | |
|----------------------------|-------------------------|-----------------|--------------------|
| Mode | Cross-correlation | Acquisition | 1 Volt / Volt |
| Length | 8192 samples (187.2 ms) | Stimulus | 1 Volt / Volt |
| Sample rate | 46 kHz (21.7 μ s) | Stimulus | |
| Concurrent pre-average | 8 | Burst MLS | |
| Auto range | Enabled | Amplitude | \pm 0.4922 Volts |
| Antialiasing filter | | Rep rate | 2.806 Hz |
| Butterworth | | Period | 16383 Samples |
| Bandwidth | 10 kHz | | 356.3 ms |
| Gain | 2 (\pm 2.5 V range) | | |
| Trigger | | | |
| Type | Stimulus Trigger | | |
| Delay | 0 samples (0 ms) | | |

Table D.4 Parameter settings of the MLSSA software for the lab system frequency response measurements

As is evident in Figures D.8 through D.12, although there are similarities in the characteristic curve of the five frequency responses, there are noticeable differences as well. This is due to many reasons, primarily the difference in early reflections off the computer monitor and desk and the different couplings of the loudspeakers to room modes due to their different locations. Despite these differences, the loudspeakers are remarkably well matched and can be used to produce reliable listening test results.

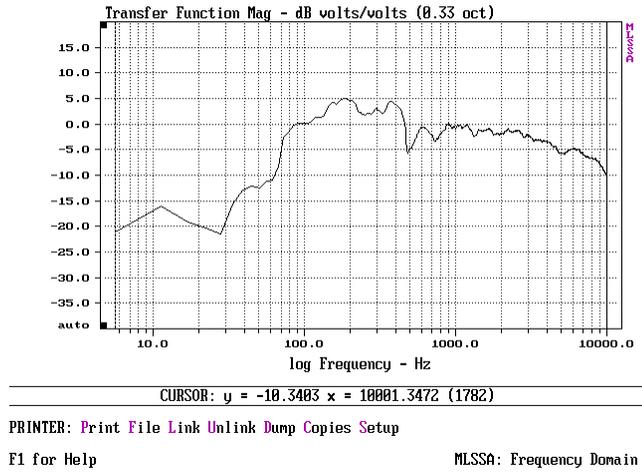


Figure D.8 One-third octave smoothed frequency response of centre loudspeaker including MARLab room response.

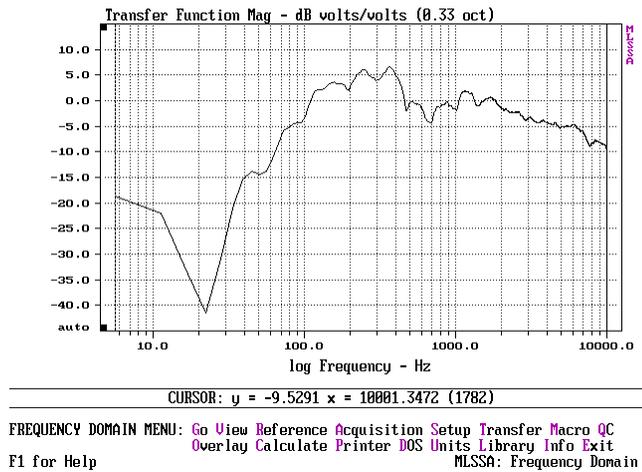


Figure D.9 One-third octave smoothed frequency response of left loudspeaker including MARLab room response.

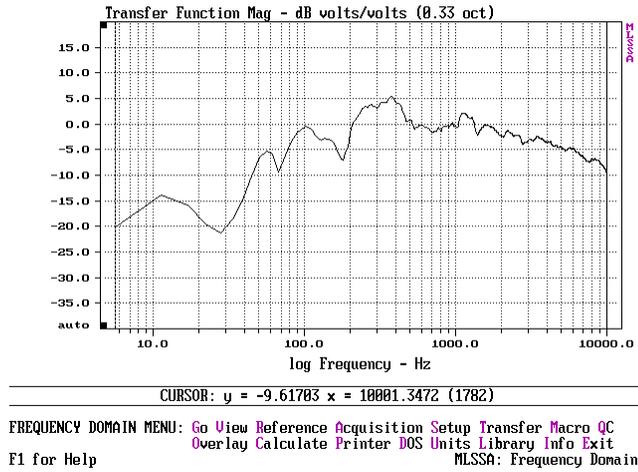


Figure D.10 One-third octave smoothed frequency response of right loudspeaker including MARLab room response.

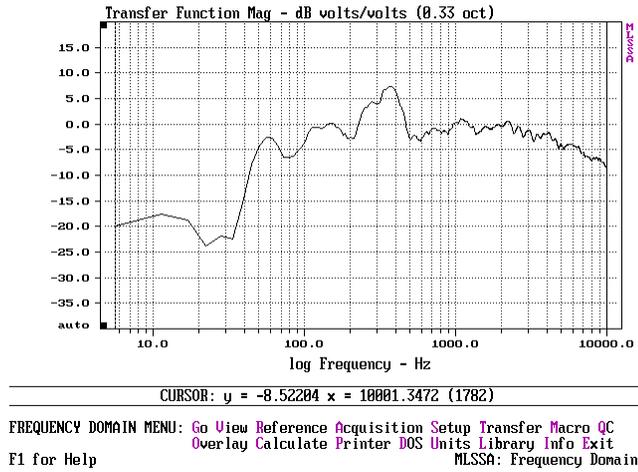


Figure D.11 One-third octave smoothed frequency response of left surround loudspeaker including MARLab room response.

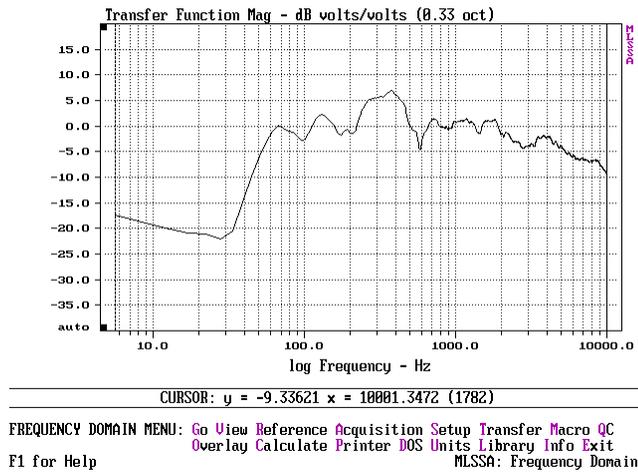


Figure D.12 One-third octave smoothed frequency response of right surround loudspeaker including MARLab room response.

Appendix E: Listening test instructions

Instructions for both tests were displayed on the computer monitor throughout the duration of each test.

E.1 Test 1

You are presented with 3 sound signals, labeled X, A and B

At least one of signals A or B is the same as X.

Please choose which of A or B exactly matches X.

The sound signals loop automatically.

You can change between the signals at any time.

You can hear each signal as often as you wish in any order by clicking on the buttons with the mouse, or by using the 1, 2, and 3 buttons on the keypad.

The indicator under the buttons tells you which signal is presently being played.

The indicator to the right of this tells you your answer. This answer is not recorded until you click on the "Next" button, or the ENTER key on the keypad.

Your responses are not timed.

If you wish to take a break, the PAUSE and PLAY buttons control the audio output.

E.2 Test 2

You are presented with a mix of two sound signals

Using the left and right arrow keys, or the arrows on screen, set the mix between the two to your most preferred position.

The sound signals loop automatically.

You may change the mix at any time.

Your response is not recorded until you click on the “NEXT” button, or the ENTER key on the far right of the keyboard.

Your responses are not timed.

If you wish to take a break, the PAUSE and PLAY buttons control the audio output.

Appendix F: Information regarding included CD-ROM

The enclosed CD-ROM contains applications named “stereo_demo” and “multichannel_demo” that allow users to hear the results of the processing method using a standard Apple Macintosh PPC with a stereo or multichannel audio output and loudspeaker configuration.

F.1 Stereo demonstration

The stereo demonstration of the system can be heard by opening the application named “stereo_demo” on the attached CD-ROM. The output of the system uses the same sound files as those used in the listening tests (see Chapter 5 for more information regarding virtual room size and microphone configuration). For this demonstration, the left front, left omnidirectional and left rear microphones are routed in parallel to the left output and the corresponding right microphones to the right output. The centre microphone output is not used.

Although intuition would dictate that a stereo microphone spacing of 2 m (as is the case in this model) is excessive and will result in a “hole in the middle” effect, this is not the result due to the interchannel temporal stability of the computer model, an effect seldom achieved in the real world.

F.1.1 Minimum system requirements

Apple Macintosh PowerPC with at least 20 MB of available RAM.

Stereo loudspeakers (equidistant from the listener and located at $\pm 30^\circ$).

F.1.2 Instructions and configuration

The left and right output channels of the Macintosh should be routed to left and right loudspeakers respectively. Alternatively, if desired, the output can be monitored over headphones.

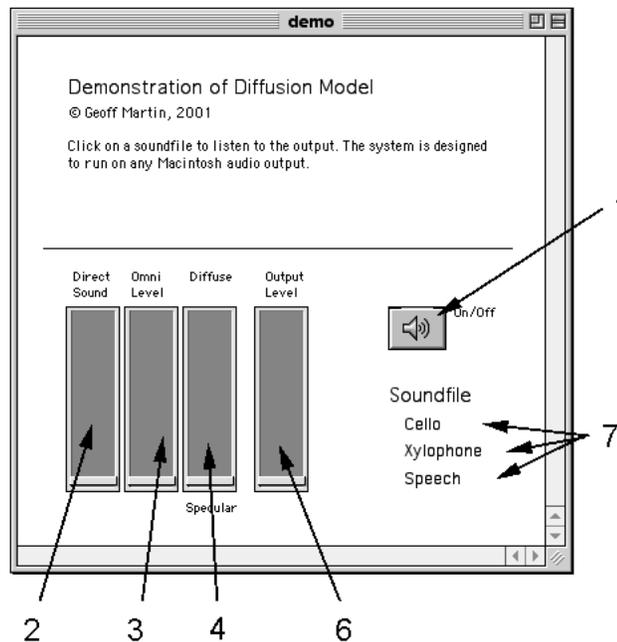


Figure F.1 Screen shot of stereo demo. Please refer to the corresponding instructions for each controller listed in Section F.2.2.

F.2 Multichannel demonstration

The stereo demonstration of the system can be heard by opening the application named “multichannel_demo” on the attached CD-ROM. In this demonstration, the left front, centre front, right front left surround and right surround cardioid microphones are all routed to their corresponding loudspeaker channel. The left omnidirectional microphone is routed to the left front and left surround loudspeakers and the right omnidirectional microphone is similarly routed to the right loudspeakers

F.2.1 Minimum system requirements

Apple Macintosh PowerPC with at least 20 MB of available RAM.

Multichannel audio output device compatible with ASIO drivers required for multichannel demonstration.

Loudspeaker configuration compatible with ITU-R BS.775-1.

Also recommended, but not required is an external multichannel digital reverberation unit

F.2.2 Instructions and configuration

The output channels of the audio device should be routed as is shown in Table F.1.

F.2.2 Instructions and configuration

| Output channel | Routing |
|----------------|--|
| 1 | External reverberation unit mono input |
| 4 | Left front loudspeaker |
| 5 | Centre front loudspeaker |
| 6 | Right front loudspeaker |
| 7 | Left surround loudspeaker |
| 8 | Right surround loudspeaker |

Table F.1 Routing for audio outputs

The reverberation unit should be set only to provide reverberation with no predelay, without direct sound or early reflections.

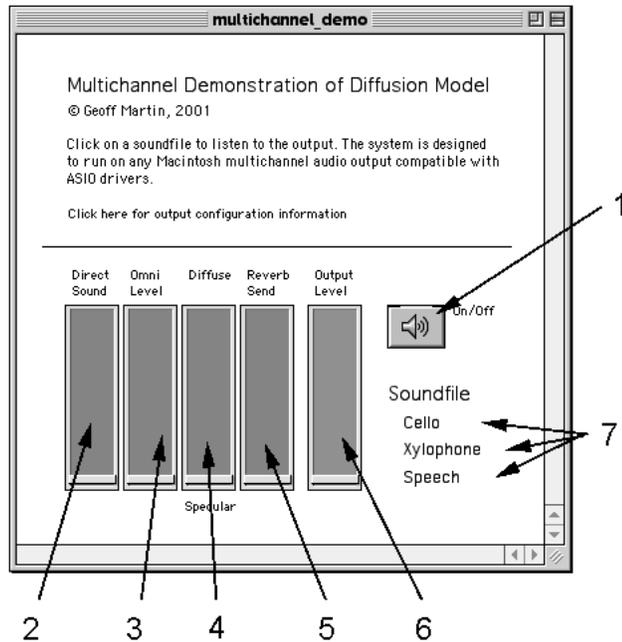


Figure F.2 Screen shot of application window for multichannel demonstration. Please refer to the corresponding instructions listed below.

- 1. On/Off** This button engages and disengages the DSP engine.

- 2. Direct Sound** This is the level controller for the direct sound arriving at the virtual microphones before the early reflections.

- 3. Omni Level** This is the level controller for the output of the omnidirectional microphones. This controller determines the level of both the direct sound and the early reflections arriving at the microphones.

- 4. Diffuse / Specular** This controls the relative balance of the diffused and specular reflection components for the four first reflections.

- 5. Reverb send** This controls the level of the output sent to channel 1 which should be routed to the external multichannel reverberation unit mono input.

- 6. Output level** This the the master output level controller for the entire system.

- 7. Soundfile** These are the three “play” buttons for the included sound files. Note that these are the same sound files used in the listening tests.

References and bibliography

Ahnert, W. (1992) "Complex Simulation of Acoustical Sound Fields," 1," *Proceedings of the 14th International Congress on Acoustics, Paper F1-10*, Beijing, 3-10 September.

Allen, J.B., & Berkley, D.A. (1979) "Image Method for Efficiently Simulating Small-Room Acoustics," *Journal of the Acoustical Society of America*, 65(4): 943-950, April.

Ando, Y. (1998) *Architectural Acoustics: Blending Sound Sources, Sound Fields, and Listeners*, AIP Press: Springer, New York.

Ando, Y., & Kato, K. (1976) "Calculations on the Sound Reflection from Periodically Uneven Surfaces of Arbitrary Profile," *Acustica*, 35: 321-329.

Ando, Y., & Noson, D. (1997) *Music and Concert Hall Acoustics*, Academic Press, London.

Ando, Y., & Schroeder, M. R. (1985) *Concert Hall Acoustics*, Springer-Verlag, Berlin.

Angus, J. A. S. (1995) "Using Modulated Phase Reflection Gratings to Achieve Specific Diffusion Characteristics," *99th Convention of the Audio Engineering Society, preprint no. 4116*, New York, 6-9 October.

Angus, J. A. S. (1999) "The Effects of Specular Versus Diffuse Reflections on the Frequency Response at the Listener," *106th Convention of the Audio Engineering Society, preprint no. 4938*, Munich, 8-11 May.

Angus, J. A. S., Marvin, A. C., Clegg, J., Dawson, J. F., & Knobloch, A. (1995) "The Effect of Acoustic Diffusers on Room Mode Decay," *99th Conference of the Audio Engineering Society*, preprint no. 4116, New York, 6-9 October.

Ballou, G., ed. (1987) *Handbook for Sound Engineers: The New Audio Cyclopedia*, Howard W. Sams & Company, Indianapolis.

Bang & Olufsen (1992) *Music for Archimedes*, CD B&O 101

Bamford, J. S., & Vanderkooy, J. (1995) "Ambisonic Sound for Us," *99th Conference of the Audio Engineering Society*, preprint no. 4138, New York, 6-9 October.

Bech, S. (1994) "Perception of Timbre of Reproduced Sound in Small Rooms: Influence of Room and Loudspeaker Position," *Journal of the Audio Engineering Society* 42: 999-1007.

Bech, S. (1995a) "Perception of Reproduced Sound: Audibility of Individual Reflections in a Complete Sound Field, II," *99th Conference of the Audio Engineering Society*, preprint no. 4093, New York, 6-9 October.

Bech, S. (1995b) "Timbral Aspects of Reproduced Sound in Small Rooms, I." *Journal of the Acoustical Society of America*, 97(3): 1717-1726.

Bech, S. (1996) "Timbral Aspects of Reproduced Sound in Small Rooms, II." *Journal of the Acoustical Society of America*, 99(6): 3539-3549.

Bech, S. (1999a) "Methods for Subjective Evaluation of Spatial Characteristics of Sound," *Proceedings of the AES 16th International Conference on Spatial Sound Reproduction*, pp. 487-504, Rovaniemi, Finland, 10-12 April.

Bech, S. (2000) *Personal communication*, December.

Bech, S., & Zacharov, N. (1999b) "Multichannel Level Alignment, Part III: The Effects of Loudspeaker Directivity and Reproduction Bandwidth," *106th Convention of the Audio Engineering Society*, preprint no. 4909, Munich, 8-11 May.

Békésy, G. von (1960) *Experiments in Hearing*, McGraw Hill, New York.

Benade, A. H. (1990) *Fundamentals of Musical Acoustics*, Dover, New York.

Beranek, L. L. (1986) *Acoustics*, Acoustical Society of America, Woodbury.

Beranek, L. L. (1988) *Acoustical Measurements.*, Acoustical Society of America, Woodbury.

Beranek, L. L. (1996) *Concert and Opera Halls: How They Sound*, Acoustical Society of America, Woodbury.

Berman, J. M. (1975) "Behaviour of Sound in a Bounded Space," *Journal of the Acoustical Society of America*, 57: 1274-1291.

Blauert, J. (1997) *Spatial Hearing: The Psychophysics of Human Sound Localization*, MIT Press, Cambridge, Revised Edition.

Bolejko, R., & Pruchnicki, P. (1998) "Sound Diffusers Based on Number Theory with Random Variations of Surface Acoustic Impedance," *104th Conference of the Audio Engineering Society*, preprint no. 4711, Amsterdam, 16-19 May.

Borish, J. (1984) "Extension of the Image Model to Arbitrary Polyhedra," *Journal of the Acoustical Society of America*, 75: 1827-1836.

Borish, J. (1986) "Diffusing Surfaces in Concert Halls: Boon or Bane?" *Journal of the Audio Engineering Society* 34(7/8): 539-545.

Borwick, J., ed. (1988) *Loudspeaker and Headphone Handbook*, Butterworths, London.

Borwick, J. (1990) *Microphones: Technology and Technique*, Focal Press, London.

Bregman, A. S. (1990) *Auditory Scene Analysis : The Perceptual Organization of Sound*, MIT Press. Cambridge.

Brüel & Kjær (1994) *Technical Documentation, Multichannel Analysis System, Type 3550, Vol. 2: Reference and Hardware*, Brüel & Kjær, Revision, December.

CATT (1998) *CATT DLL Directivity Interface (DDI) v1.0: DDI White Paper; Rev. 0*.

Cherry, E. C. (1953) "Some experiments on the recognition of speech with one and with two ears," *Journal of the Acoustical Society of America*, 25: 975-979.

Christensen, K. B. (1992) "The Application of Digital Signal Processing to Large Scale Simulation of Room Acoustics Part II: Frequency Response Modeling and Optimization Software for a Multichannel DSP Engine," *Journal of the Audio Engineering Society* 40: 260-276.

Christensen, K. B., & Lund T. (1999) "Room simulation for multichannel film and music" *107th Convention of the Audio Engineering Society, preprint no. 4993*, New York, 24-27 September.

Coleman, P. D. (1962) "Failure to Localize the Source Distance of an Unfamiliar Sound," *Journal of the Acoustical Society of America*, 34: 345-346.

Coleman, P. D. (1963) "An Analysis of Cues to Auditory Depth Perception in Free Space," *Psychological Bulletin*, 60: 302-315.

Cox, T. J. (1995) "Diffusion Parameters for Baffled Diffusers," *99th Conference of the Audio Engineering Society, preprint no. 4115*, New York, 6-9 October.

Crocker, M. J., ed. (1998) *Handbook of Acoustics*, John Wiley & Sons, New York.

D'Antonio, P. (1995) "Two Decades of Diffusor Design and Development," *99th Conference of the Audio Engineering Society, preprint no. 4114*, New York, 6-9 October.

D'Antonio, P., & Konnert., J. H. (1984a) "The RPG Reflection Phase Grating Acoustical Diffusor: Experimental Measurements," *76th Conference of the Audio Engineering Society, preprint no. 2158*, New York, 8-11 October.

D'Antonio, P., & Konnert, J. H. (1984b) "The Reflection Phase Grating Diffusor: Design Theory and Application," *Journal of the Audio Engineering Society*, 32(4): 228-238, April.

Dalenbäck, B.-I. (1992) "Room Acoustics Prediction and Auralization Based on Extended Image Source Model," Licentiate thesis, rep. F92-03, Department of Applied Acoustics, Chalmers University of Technology, Gothenburg, Sweden.

Dalenbäck, B.-I. (1996) "Room Acoustic Prediction Based on a Unified Treatment of Diffuse and Specular Reflection," *Journal of the Acoustical Society of America*, 100(2): 899-909, August.

Dalenbäck, B.-I. (2001) *Personal communication*, February.

Dalenbäck, B.-I., Kleiner, M., & Svensson, P. (1994) "A Macroscopic View of Diffuse Reflection," *Journal of the Audio Engineering Society*, 42(10): 793-805, October.

Dalenbäck, B.-I., Svensson, P., & Kleiner, M. (1992) "Prediction and Auralization Based on a Combined Image Source / Ray-Model," *Proceedings of the 14th International Congress on Acoustics, Paper F2-7*, Beijing, 3-10 September.

Depalle, P. (1995) "Algorithmes de transformation de Fourier rapide," École supérieure d'électricité, 10612.

Edwards, D., & Hamson, M. (1989) *Guide to Mathematical Modelling*, CRC Press, Boca Raton, Florida.

Evans, M. J. (1998a) "Obtaining Accurate Responses in Directional Listening Tests," *104th Conference of the Audio Engineering Society, preprint no. 4730*, Amsterdam, 16-19 May.

Evans, M. J., Tew, A. I.; & Angus, J. A. S. (1998b) "Relative Spatialisation Performance of Ambisonic and Transaural Speech," *104th Conference of the Audio Engineering Society, preprint no. 4711*, Amsterdam, 16-19 May.

Eyring, C. F. (1930) "Reverberation Time in 'Dead' Rooms," *Journal of the Acoustical Society of America*, 1: 217-241.

Fahy, F. J. (1989) *Sound intensity*, Elsevier Applied Science, London.

Flanagan, S., & Taylor, V. (1999) "Investigation into the Relationship Between Subjective Loudness and Auditory Distance Perception," *107th Conference of the Audio Engineering Society, preprint no. 5049*, New York, 24-27 September.

Fukada, A., Tsujimoto, K., & Akita, S. (1997) "Microphone Techniques for Ambient Sound on a Music Recording," *103th Conference of the Audio Engineering Society, preprint no. 4540*, New York, 26-29 September.

Gardner, W. G. (1992a) "A Realtime Multichannel Room Simulator," *Journal of the Acoustical Society of America*, 92, 2395.

Gardner, W. G. (1992b) "The Virtual Acoustic Room," Master's Thesis, Music and Cognition Group, Media Lab, MIT, Cambridge.

Gardner, W. G., & Greisinger, D. (1994) "Reverberation Level Matching Experiments," *Proceedings of the Sabine Centennial Symposium*, Acoustical Society of America, 263-266.

Gibbs, B. M., & Jones, D. K. (1972) "A Simple Image Method for Calculating the Distribution of Sound Pressure Levels Within an Enclosure," *Acustica*, 26: 24-32.

Glanz, J. (2000) "Art + Science = Beautiful Music" *New York Times*, New York, April 18.

Golomb, S. W. (1967) *Shift Register Sequences*, Holden Day, San Francisco.

Gourad, H. (1971) "Continuous Shading of Curved Surfaces," *IEEE Transactions on Computers*, C-20(6): 623-629.

Hall, D. E. (1980) *Musical Acoustics: An Introduction*, Wadsworth Publishing, Belmont.

Hann, C. N., & Fricke, F. R. (1993) "Surface Diffusivity as a Measure of Acoustic Quality of Concert Halls," *Proceedings of the Conference of the Australia and New Zealand Architectural Science Association*, Sydney, 81-90.

Hidaka, T, Beranek, L., & Okano, T. (1995) "Interaural Cross Correlation (IACC), Lateral Fraction (LF) and Sound Energy Level (G) as Partial Measures of Acoustical Quality in Concert Halls," *Journal of the Acoustical Society of America*, 98:988-1007.

Hodgson, M. (1992) "Evidence of Diffuse Surface Reflections in Rooms," *Journal of the Acoustical Society of America* 89: 765-771.

Hynninen, J., & Zacharov, N. (1999) "GuineaPig – A Generic Subjective Test System for Multichannel Audio," *106th Convention of the Audio Engineering Society, preprint no. 4871*, Munich, 8-11 May.

Isaacs, A., ed. (1990) *A Concise Dictionary of Physics*, Oxford University Press, New York, New Edition.

ITU (1994) *ITU-R BS.775-1 Recommendation: Multichannel Stereophonic Sound System with and without Accompanying Picture*, International Telecommunication Union, Geneva.

Jordan, E. J. (1963) *Loudspeakers*, Focal Press, London.

Jot, J.-M., & Chaigne, A. (1991) "Digital Delay Networks for Designing Artificial Reverberators," *90th Convention of the Audio Engineering Society, preprint no. 3030*, Paris, 19-22 February.

Jot, J.-M., Cerveau, L., & Warusfel, O. (1997) "Analysis and Synthesis of Room Reverberation Based on a Statistical Time-Frequency Model," *103th Conference of the Audio Engineering Society, preprint no. 4629*, New York, 26-29 September.

Karplus, K., & Strong, A. (1983) "Digital Synthesis of Plucked-String and Drum Timbres," *Computer Music Journal*, 7(2): 42:55.

Kaup, A., Khoury, S., Freed, A., & Wessel, D. (1999) "Volumetric Modeling of Acoustic Fields in CNMAT's Sound Spatialization Theatre," *Proceedings of the International Computer Music Conference*, Beijing, 488-491, 22-28 October.

Kinsler, L. E., & Frey, A.R. (1962) *Fundamentals of Acoustics*, John Wiley & Sons, New York, 2nd edition.

Kinsler, L. E., Frey, A. R., Coppens, A. B., & Sanders, J. V. (1982) *Fundamentals of Acoustics* John Wiley & Sons, New York, 3rd edition.

Kleiner, M., Dalenbäck, B.-I., & Svensson, P. (1993) "Auralization – An Overview," *Journal of the Audio Engineering Society*, 41(11): 861-875, November.

Kurozumi, K., & Ohgushi, K. (1983) "The Relationship Between the Cross-correlation Coefficient of Two-channel Acoustic Signals and Sound Image Quality," *Journal of the Acoustical Society of America*, 74: 1726-1733.

Kuttruff, K. H. (1991) *Room Acoustics*, Elsevier Science Publishers, Essex.

Kuttruff, K. H. (1992) "Simulation Models for Auralization in Room Acoustics," *Proceedings of the 14th International Congress on Acoustics, Paper F2-2*, Beijing, 3-10 September.

Kuttruff, K. H., & Straßen, Th. (1980) "On the Dependence of Reverberation Time on the 'Wall Diffusion' and on Room Shape," *Acustica*, 45: 246-255.

Laird, J., Masri, P., & Canagarajah, C. N. (1998) "Efficient and Accurate Synthesis of Circular Membranes using Digital Waveguides," *Proceedings of the IEE Colloquium on Audio and Music Technology: The Challenge of Creative DSP*, 1-6 December.

Laird, J., Masri, P., & Canagarajah, C. N. (1999) "Modelling Diffusion at the Boundary of a Digital Waveguide Mesh," *Proceedings of the International Computer Music Conference*, Beijing, 492-495, 22-28 October.

Lehnert, H., & Blauert, J. (1992) "Principles of Binaural Room Simulation." *Applied Acoustics*, 36: 259-291.

Lewers, T. (1993) "A Combined Beam Tracing and Radiant Exchange Computer Model of Room Acoustics," *Applied Acoustics*, 38: 161 – 178.

Lexicon (1993) *Owner's Manual, Lexicon 480L Digital Effects System*, Lexicon, Waltham.

Lund, T. (2000) "Enhanced Localization in 5.1 Production," *109th Convention of the Audio Engineering Society*, preprint no. 5243, Los Angeles, 22-25 September.

Maercke, D. van (1986) "Simulation of Sound Fields in Time and Frequency Domain Using a Geometrical Model," *Proceedings of the 12th International Congress on Acoustics, Paper E11-7*, Toronto, 24-31 July.

Malcurt, C., & Jullien, J.-P. (1986) "Sound Field Prediction in a Variable Acoustic Hall," *Proceedings of the 12th International Congress on Acoustics, Paper E5-1*, Toronto, 24-31 July.

Malham, D. G. (1999) "Higher Order Ambisonic Systems for the Spatialisation of Sound," *Proceedings of the International Computer Music Conference*, Beijing, 492-495, 22-28 October.

Martens, W. L. (1999) "The Impact of Decorrelated Low-Frequency Reproduction on Auditory Spatial Imagery: Are Two Subwoofers Better than One?" *Proceedings of the AES 16th International Conference on Spatial Sound Reproduction*, pp. 67-77, Rovaniemi, Finland, 10-12 April.

Martin, G. (1996) "För en bättre förståelse av stereomikrofonteknik," *Musik & Ljudteknik*, 37(4), 1996 & 38(1), 1997.

Martin, G., Woszczyk, W., Corey, J., & Quesnel, R. (1999a) "Controlling Phantom Image Focus in a Multichannel Reproduction System," *107th Conference of the Audio Engineering Society, preprint no. 4996*, New York, 24-27 September.

Martin, G., Woszczyk, W., Corey, J., & Quesnel, R. (1999b) "Sound Source Localization in a Five-Channel Surround Sound Reproduction System," *107th Conference of the Audio Engineering Society, preprint no. 4994*, New York, 24-27 September.

MathWorks (1998) *MATLAB: Signal Processing Toolbox for use with MATLAB*, The MathWorks, Natick, Revised for MATLAB 5.2.

McClellan, J. H., Schafer, R. W., Yoder, M. A. (1998) *DSP First : A Multimedia Approach*, Prentice Hall, Saddle River.

Mehta, M. L., & Mulholland, K. A., (1976) "Effect of Non-Uniform Distribution of Absorption on Reverberation Time," *Journal of Sound and Vibration*, 46(2) 209-224.

Mershon, D. H., & Bowers, J. N. (1979) "Absolute and Relative Cues for the Auditory Perception of Egocentric Distance," *Perception* 8: 311-322

Mershon, D. H., Ballenger, W. L., Little, A. D., McMutry, P. L., & Buchanan, J. L. (1989) "Effects of Room Reflectance and Background Noise on Perceived Auditory Distance." *Perception*, 18: 403-416.

Metzler, B. (1993) *Audio Measurement Handbook*, Audio Precision, Beaverton.

Meyer, J. (1975) *Instrumentenbau*, 29(2).

Moore, B. C. J. (1997) *An Introduction to the Psychology of Hearing*, Academic Press, San Diego, 4th Edition.

Moore, F. R. (1983) "A General Model for Spatial Processing of Sounds," *Computer Music Journal*, 7(3): 6-15.

- Moorer, J. A. (1979) "About This Reverberation Business." *Computer Music Journal*, 3(2): 13-28.
- Nakagawa, K., Miyajima, T., & Tahara, Y. (1993) "An Improved Geometrical Sound Field Analysis in Rooms Using Scattered Sound and an Audible Room Acoustic Simulator," *Applied Acoustics*, 38: 115-129.
- Naylor, G. (1993) "Treatment of Early and Late Reflections in a Hybrid Model for Room Acoustics," *Journal of the Acoustical Society of America (Abstracts)*, 92: 145-159.
- Neilsen, S., H. (1991) "Distance Perception in Hearing," Doctoral thesis, Aalborg University, Aalborg, April.
- Neter, J., Wasserman, W., & Whitmore, G. A. (1993) *Applied Statistics*, Allyn and Bacon, Boston, 4th Edition.
- Nuttgens P. (1997) *The Story of Architecture*, Phaidon Press, London, 2nd Edition.
- Olive, S. E., & Toole, F. E. (1989) "The Detection of Reflections in Typical Rooms." *Journal of the Audio Engineering Society*, 37: 539-553.
- Olson, H. F. (1957) *Acoustical engineering*, Van Nostrand, Princeton.
- Peterson, P. M. (1986) "Simulating the Response of Multiple Microphones to a Single Acoustic Source in a Reverberant Room," *Journal of the Acoustical Society of America*, 80(5): 1527-1529, November.

Phong, B. T. (1975) "Illumination for Computer Generated Pictures," *Communications of the Association for Computing Machinery*, 18(6): 311-317, June.

Quesnel, R., Woszczyk, W., Corey, J., & Martin, G. (1999) "A Computer System for Investigating and Building Synthetic Auditory Spaces – Part 1," *107th Conference of the Audio Engineering Society, preprint no. 4992*, New York, 24-27 September.

Rabiner, L. R., & Juang, B. H. (1986) "An Introduction to Hidden Markov Models," *IEEE ASSP Magazine*, 4-16, January.

Rayleigh, J. W. S. (1945) *The Theory of Sound*, Dover, New York.

Rietschote, H. F. van, Houtgast, T., & Steeneken, H. J. M. (1981) "Predicting Speech Intelligibility in Rooms from the Modulation Transfer Function IV: A Ray-Tracing Computer Model," *Acustica*, 49:245-252.

Rife, D. D. (2000) *MLSSA: Maximum-Length Sequence System Analyzer, Reference Manual Version 10WA*, DRA Laboratories, Sarasota.

Rindel, J. H. (1997) "Computer Simulation Techniques for Acoustical Design of Rooms - How to Treat Reflections in Sound Field Simulations," *Proceedings of ASVA 97*, 201-208, Tokyo, 2-4 April.

Rindel, J. H., & Naylor, G. (1992) "Computer Modelling of Sound Fields in Rooms – The State of the Art and Outlook to the Future," *Proceedings of the 14th International Congress on Acoustics, Paper F2-1*, Beijing, 3-10 September.

Roads, C., ed. (1989) *The Music Machine: Selected Readings from Computer Music Journal*, MIT Press, Cambridge.

Roads, C., ed. (1996) *The Computer Music Tutorial*, MIT Press, Cambridge.

Roads, C., & Strawn, J., ed. (1985) *Foundations of Computer Music*, MIT Press, Cambridge.

Rubak, P., & Johansen, L. G. (1998) "Artificial Reverberation Based on Pseudo-Random Impulse Response," *104th Conference of the Audio Engineering Society*, preprint no. 4725, Amsterdam, 16-19 May.

Rubak, P., & Johansen, L. G. (1999) "Artificial Reverberation Based on Pseudo-Random Impulse Response: II," *106th Convention of the Audio Engineering Society*, preprint no. 4900, Munich, 8-11 May.

Savioja, L. (1999) "Modeling Techniques for Virtual Acoustics," Doctoral thesis, Helsinki University of Technology, Telecommunications Software and Multimedia Laboratory, Report TML-A3.

Schroeder, M. R. (1961) "Improved Quasi-Stereophony and Colorless Artificial Reverberation." *Journal of the Acoustical Society of America*, 33: 1061.

Schroeder, M. R. (1962) "Natural Sounding Reverberation," *Journal of the Audio Engineering Society*, 10(3): 219-223.

Schroeder, M. R. (1970) "Digital Simulation of Sound Transmission in Reverberant Spaces" *Journal of the Acoustical Society of America*, 47(2): 424-431.

Schroeder, M. R. (1975) "Diffuse Reflection by Maximum-Length Sequences," *Journal of the Acoustical Society of America*, 57(1): 149-150.

Schroeder, M. R. (1979) "Binaural Dissimilarity and Optimum Ceilings for Concert Halls: More Lateral Sound Diffusion," *Journal of the Acoustical Society of America*, 65(4): 958-963.

Schroeder, M. R., Gottlob, D., & Siebrasse, K.F. (1974) "Comparative Study of European Concert Halls: Correlation of Subjective Preference with Geometric and Acoustic Parameters," *Journal of the Acoustical Society of America*, 56(4): 1195-1201.

Sollich, P. (2001) "Markov equation, Markov process," <<http://www.mth.kcl.ac.uk/~psollich/StatMechNN/node10.html>>, 20 February.

Siegel, A. F., & Morgan, C. J. (1996) *Statistics and Data Analysis: An Introduction*, John Wiley & Sons, New York, 2nd Edition.

Stautner, J., & Puckette, M. (1982) "Designing Multi-channel Reverberators," *Computer Music Journal*, 6(1): 52-65.

Steiglitz, K. (1996) *A DSP Primer : With Applications to Digital Audio and Computer Music*, Addison-Wesley, Menlo Park.

Stone, H., & Sidel, J. L. (1993) *Sensory Evaluation Practices*, Academic Press, San Diego, 2nd Edition.

Suokuisma, P., Zacharov, N., & Bech, S. (1998) "Multichannel Level Alignment, Part I: Signals and Method," *105th Convention of the Audio Engineering Society, preprint no. 4815*, San Francisco, 26-29 September.

Tan Kate, W. R. Th. (1992) "Analysis of the Diffusor," *Proceedings of the 14th International Congress on Acoustics, Paper F8-6*, Beijing, 3-10 September.

Theile, G., & Plenge, G. (1977) "Localization of Lateral Phantom Sources," *Journal of the Audio Engineering Society*, 25(4): 196-200, April.

Van Duyne, S. A., & Smith J. O. (1993) "Physical Modeling with the 2-D Digital Waveguide Mesh," *Proceedings of the 1993 International Computer Music Conference*, Tokyo, 40-47.

Vian, J.-P., & Martin, J. (1992) "Binaural Room Acoustics Simulation: Practical Uses and Applications," *Applied Acoustics*, 36: 293-305.

Vörländer, M. (1989) "Simulation of the Transient and Steady-State Sound Propagation in Rooms Using a New Combined Ray-Tracing / Image-Source Algorithm," *Journal of the Acoustical Society of America*, 86: 72-178, July.

Wallach, H. (1949) "The Precedence Effect in Sound Localization," *American Psychology*, 62: 315-336.

Warusfel, O., Denayrou, R., & Jullien, J.-P. (1992) "Validation of a Computer Model Environment for Room Acoustics Prediction," *Proceedings of the 14th International Congress on Acoustics, Paper F6-5*, Beijing, 3-10 September.

Watt, A. H. (1989) *Fundamentals of Three-dimensional Computer Graphics*, Addison-Wesley, Wokingham.

Woram, J. M. (1989) *Sound Recording Handbook*, Howard W. Sams & Company, Indianapolis.

Zacharov, N., Bech, S., & Suokuisma, P. (1998) "Multichannel Level Alignment, Part II: The Influence of Signals and Loudspeaker Placement," *105th Convention of the Audio Engineering Society, preprint no. 4816*, San Francisco, 26-29 September, 1998.

Zölzer, U. (1997) *Digital Audio Signal Processing*, John Wiley & Sons, Chichester.

Zwicker, E., & Fastl, H. (1999) *Psychoacoustics: Facts and Models*, Springer, Berlin.