In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Mechanical Engineering

The University of British Columbia
Vancouver, Canada

Date October 15, 1997.
- ABSTRACT -

Sound fields in rooms, including the complex interactions between propagating sound waves and the room surfaces, are predicted using various approaches. These approaches are validated either physically - by comparing the measured and predicted sound fields - or subjectively. Subjective evaluation of the predictions can be performed using auralization, which consists of simulating an acoustical environment for binaural presentation to a listener. Inaccuracies that result from the different sound-field prediction methods result in imperfect auralization. It is of interest to understand the perceptual consequences of these inaccuracies and compensate for them with accurate sound-field prediction models. The aim of this research project was to develop accurate sound-field prediction models for use in auralization systems. To achieve this objective, two phases of work were undertaken: (1) acoustical signals were created, using simplified sound-field techniques, and presented to listeners using a commercial auralization system in order to test sound-localization ability; (2) a sound-prediction algorithm was derived and used in the development of an improved room-prediction model. With respect to the work using the commercial auralization system, modifications were made to the system to mimic higher-order reflections in the room, as well as to account for varying surface absorption in each octave band. Subjects' localization abilities were evaluated using the sound-field simulations and in a real room to validate the sound-field simulations. Inconclusive results from the localization tests led to the development of an improved sound-field prediction model based on acoustical radiosity. The radiosity model was validated experimentally in full-scale and scale-model rooms with the help of other prediction models. A combined model - based on the method of images and radiosity - was developed and validated in several rooms. An inherent attribute of the radiosity approach which makes it unsuitable for auralization was identified.
# TABLE OF CONTENTS

Abstract

Table of Contents

List of Tables

List of Figures

Acknowledgement

1 Introduction

  1.1 Room Simulation and Auralization .............................................. 1
  1.2 Architectural Acoustics and Sound-Field Prediction ......................... 4
  1.3 Development of Accurate Sound-Field Prediction Models ................. 7
  1.4 Thesis Organization ............................................................ 7

Section I: Room Sound-Field Prediction and Auralization for Hearing Research

2 Technical Background

  2.1 Localization in Real Environments ............................................ 9
    2.1.1 Free-Field Auditory Localization ...................................... 11
    2.1.2 Reverberant-Field Auditory Localization ............................. 13
  2.2 Localization in Virtual Environments ....................................... 14
    2.2.1 Head-Related Impulse Response ....................................... 15
    2.2.2 Replay Transducer ................................................................ 17
    2.2.3 Sound Source ....................................................................... 18
    2.2.4 Room Impulse Response (RIR) .............................................. 19
  2.3 Detailed Research Objectives ..................................................... 23
3 Auralization Using the Tucker-Davis System

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Hardware and Software Features/Limitations of the Auralization System</td>
<td>25</td>
</tr>
<tr>
<td>3.2</td>
<td>Localization Conditions</td>
<td>28</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Sound-Field Simulation</td>
<td>28</td>
</tr>
<tr>
<td>3.2.2</td>
<td>Non-Individualized HRTF</td>
<td>33</td>
</tr>
<tr>
<td>3.2.3</td>
<td>Feedback Condition</td>
<td>35</td>
</tr>
</tbody>
</table>

4 Auralization System Input Data

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Measurement of the Room Impulse Response</td>
<td>36</td>
</tr>
<tr>
<td>4.1.1</td>
<td>Room Surface-Absorption Coefficients</td>
<td>37</td>
</tr>
<tr>
<td>4.1.2</td>
<td>Room Reverberation Times</td>
<td>40</td>
</tr>
<tr>
<td>4.2</td>
<td>HRTFs</td>
<td>42</td>
</tr>
<tr>
<td>4.3</td>
<td>Headphone-Ear Replay-Compensation Filter</td>
<td>44</td>
</tr>
</tbody>
</table>

5 Validation of the Room Simulation

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Experimental Arrangement</td>
<td>47</td>
</tr>
<tr>
<td>5.2</td>
<td>Source Signals and Test Procedure</td>
<td>49</td>
</tr>
<tr>
<td>5.3</td>
<td>Results and Discussion of the Localization Tests</td>
<td>50</td>
</tr>
<tr>
<td>5.4</td>
<td>Consideration of a Different Prediction Algorithm</td>
<td>53</td>
</tr>
</tbody>
</table>

Section II: Development of an Improved Prediction Algorithm

6 Technical Background

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1</td>
<td>Literature Review and Critique</td>
<td>55</td>
</tr>
<tr>
<td>6.2</td>
<td>Detailed Research Objectives</td>
<td>57</td>
</tr>
<tr>
<td>6.3</td>
<td>Theoretical Development</td>
<td>57</td>
</tr>
<tr>
<td>6.4</td>
<td>Digital-Signal-Processing (DSP) Issues</td>
<td>62</td>
</tr>
<tr>
<td>6.4.1</td>
<td>Sampling-Frequency Determination</td>
<td>63</td>
</tr>
</tbody>
</table>
- LIST OF TABLES -

4.1 Absorption Coefficients for the Variable-Acoustic Test Room Surfaces .......... 38
4.2 Reverberation Times of the Furnished Test Room in Octave Bands ............... 40
4.3 MLSSA Calculation of Reverberation Times for each Octave Band
    from the Impulse Response of Figure 4.3. ........................................... 41

6.1 Description of Radiosity M-files in the First Module of the Program ............... 72

7.1 Digital Equalizer Settings in 1/3 Octave Bands for Full-Scale
    and Model Rooms ................................................................................. 76
7.2 Sound Power of the Full-Scale and Model Sources, in Octave Bands .............. 77
7.3 Air Absorption Exponents used in Predictions of the Environmental Room ....... 78
7.4 Air Absorption Exponents used in Predictions for Hebb 12 ............................. 80
7.5 Air Absorption Exponents used in Predictions (Scale-Model Values) .............. 82

8.1 Radiosity Prediction versus Diffuse-Field Theory Results for
    a Source at [1, 1, 1] = [x, y, z] .................................................................. 88
- LIST OF FIGURES -

1.1 Schematic Describing the Propagation of Sound from a Source to Listener's Eardrum in Real and Simulated Rooms ......................................................... 2
1.2 Specular, Partly Diffuse, and Diffuse Reflections ............................................. 5
1.3 Illustration of the Room Impulse Response ...................................................... 6

2.1 Coordinate System For Sound Localization ..................................................... 10

3.1 Auralization System Hardware Schematic...................................................... 26
3.2 AutoRoute Diagram for the "Direct Sound Only" Condition ............................. 29
3.3 AutoRoute Diagram for the "Direct Sound Plus First-Order Reflections" Condition ........................................................... 30
3.4 AutoRoute Schematic for the Reverberant Tail ................................................ 33
3.5 Magnitude Response of HRTFs for a Source at 0° : (a) Right Ear of SJX, (b) Left Ear of SJX .......................................................... 34

4.1 Measured Room Impulse Response of the Test Room ................................... 37
4.2 (a) Ceiling, (b) Floor, and (c) Wall Filters ...................................................... 39
4.3 Room Impulse Response of the Simulated Test Room ................................... 41
4.4 Horizontal Plane HRTFs for an Azimuthal Angle of 90°: (a) SJX, (b) SOS, (c) SOU .......................................................... 43
4.5 Set Up for Measuring the Headphone-Ear Replay-Compensation Filter ............ 44
4.6 HD 265 and KUNOV Ear Compensation Filter .............................................. 45

5.1 Speaker and Subject Positions for the Real Test Room ................................... 48
5.2 Localization Test Results in the Real Room: (a) Without Feedback (b) With Feedback ................................................................. 51

5.3 Localization Test Results in the Virtual Room Using HRTF from SOU without Feedback: (a) Direct Sound (b) Direct Sound with First-Order Reflections ...................................................................................... 52

5.4 Localization Test Results in the Virtual Room with Direct Sound and First-Order Reflections without Feedback: (a) Using SOS's HRTF (b) Using SJX's HRTF (c) Using SOU's HRTF ............................................................................................................ 52

6.1 Test Chamber Containing an Omnidirectional Loudspeaker and a Microphone Receiver with Surfaces Divided into Patches .......................................................... 58

6.2 Patch j Receiving Intensity $\Phi_j$ From Patch i ....................................................................................................................... 58

6.3 Geometry for Form-Factor Derivation (extracted from Goral et al., 1984 [45]) .................................................................................................................................................. 59

6.4 Sound Propagation from Patch i to Patch j (Extracted from Shi et al., 1993 [46]) .............................................................................................................................................. 61

6.5 Frequency Spectrum of a Signal at Different Times .................................................................................................................. 64

6.6 Planar Sound Source and Observer (Sound-Observer Distance = c) ......................................................................................... 66

6.7 Program Flow Chart for Predicting the RIR Using Radiosity and Method-of Images Approaches .............................................................................................................. 70

7.1 Measurement System for Calibrating the Power of the Source ......................................................................................... 74

7.2 Source and Receiver Positions in the Environmental Room: (a) Receiver at (2 m, 2.36 m, 1.29 m), (b) Receiver at (2 m, 2.36 m, 1.5 m), and (c) Receiver at (1 m, 2.36 m, 1.29 m). Source position in all three cases was (2 m, 4.86 m, 1.29 m) .................................................................................................................. 78

7.3 Dimensions of Hebb 12 ................................................................................................................................................... 80

7.4 Absorption Coefficients for Hebb 12’s Surfaces ............................................................................................................. 80
7.5 Source and Receiver Positions in Hebb 12: (a) Receiver at (5.3 m, 3.9 m, 1.3 m), (b) Receiver at (12 m, 3.9 m, 1.3 m), and (c) Receiver at (11.7 m, 5.8 m, 1.3 m). Source position in all three cases was (1.5 m, 3.9 m, 1.3 m)…………………………… 81

7.6 Dimensions of the Scale-Model Room ……………………………………………………………………………………………………………… 82

7.7 Average Absorption Coefficients of Scale-Model Room’s Surfaces…………………….. 82

7.8 Source and Receiver Positions in Scale Model Rooms: (a) Receiver at (1.13 m, 0.875 m, 0.10 m), (b) Receiver at (1.9 m, 0.875 m, 0.10 m) Source position in both cases was (0.63 m, 0.875 m, 1.40 m)……………………………………… 83

8.1 (a) Octave-Band Reverberation Times and (b) Steady-State Sound-Pressure Levels at Receiver Position 3 for the Environmental Room; ... 96 patches, --- 150 patches, ___ 96 patches with longer truncation times…………………………………… 85

8.2 Comparison of Sound Propagation Between a Source and a Receiver for Different Patch Sizes …………………………………………………………………………………… 86

8.3 Echogram Predicted Using Radiosity for a Receiver at [2.9, 2.9, 2.9] ……………… 88

8.4 Predicted Reverberation Times from Various Models for the Three Receiver Positions in the Environmental Room: (a) Receiver Position 1, (b) Receiver Position 2, (c) Receiver Position 3; *-- radiosity model, x-- diffuse ray-tracing model, o-- specular ray-tracing model, and +-- method-of-images model ………………………………………………………………………………… 91

8.5 Predicted L_p from Various Models for the Three Receiver Positions in the Environmental Room: (a) Receiver Position 1, (b) Receiver Position 2, (c) Receiver Position 3; *-- radiosity model, x-- diffuse ray-tracing model, o-- specular ray-tracing model, and +-- method-of-images model ………………………………………………………………………………… 92

8.6 1000 Hz Echogram Predicted using Various Models: (a) Radiosity, (b) Diffuse Ray-Tracing, (c) Specular Ray Tracing, (d) Method of Images….. 94

8.7 Predicted RIR using Various Models: (a) Radiosity, (b) Diffuse Ray-Tracing, (c) Specular Ray Tracing, (d) Method of Images……………………………………………………………………………… 95
8.8 (a) Impulse Response of the Environmental Room, (b) Energy Decay-Curve, as Measured at Receiver Position 3 ................................................................. 96
8.9 (a) Impulse Response of the Hebb 12, (b) Energy-Decay Curve for Hebb 12, as Measured at Receiver Position 3 ................................................................. 97
8.10 (a) Impulse Response the 10 m height Scale Model Room, (b) Energy-Decay Curve for the 10 m height Scale Model Room, as Measured at Receiver Position 2 ...... 97
8.11 (a) Impulse Response the 5 m height Scale Model Room, (b) Energy-Decay Curve for the 5 m height Scale Model Room, as Measured at Receiver Position 2 ...... 98
8.12 Echograms for the Environmental Room: (a) Predicted Response for the 1000 Hz Octave Band, (b) Measured Wide-Band Response ........................................ 98
8.13 Echograms for Hebb 12: (a) Predicted Response for the 1000 Hz Octave Band, (b) Measured Wide-Band Response ........................................ 99
8.14 Echograms for the 10 m High Scale-Model Room: (a) Predicted Response for the 8000 Hz Octave Band, (b) Measured Wide-Band Response ........................................ 99
8.15 Echograms for the 5 m High Scale-Model Room: (a) Predicted Response for the 8000 Hz Octave Band, (b) Measured Wide-Band Response ........................................ 100
8.16 Percent Deviation of $T_{60}$ for all Four Rooms as a Function of Frequency:
(a) Environmental Room, (b) Hebb 12, (c) Scale-Model Room with 10 m Height, (d) Scale Model Room with 5 m Height; $^*$ -- Receiver Position 1, $^\times$ -- Receiver Position 2, $^\circ$ -- Receiver Position 3 ................................................................. 101
8.17 Difference Between Measured and Predicted $L_p$ for all Four Rooms as a Function of Frequency: (a) Environmental Room, (b) Hebb 12, (c) Scale-Model Room with 10 m Height, (d) Scale Model Room with 5 m Height; $^*$ -- Receiver Position 1, $^\times$ -- Receiver Position 2, $^\circ$ -- Receiver Position 3 ................................................................. 102
- ACKNOWLEDGEMENTS -

First and foremost, I would like to thank my thesis supervisor, Dr. Murray Hodgson, for his ongoing support, enthusiasm and assistance throughout my research. His expertise in acoustics was invaluable. I must also thank Mr. Ian Ashdown for contributing his expertise in radiosity and providing assistance wherever possible. Ian has been instrumental in the development of the acoustical radiosity model.

I would like to also express a word of gratitude to the graduate students in the acoustics group of Mechanical Engineering - especially Nelson Heerema - for their informal and helpful consultation. Moreover, I would like to express my appreciation to the Computer Systems Administrator for Mechanical Engineering, Mr. Alan Steeves, for his assistance on computing matters. A graduate student in Electrical Engineering, Dan Lisogurski, deserves special mention for his assistance with UNIX and his patience during my learning experience.

In addition, several individuals in the School of Audiology and Speech Sciences have been very helpful in performing localization tests. Dr. Kathy Pichora-Fuller and Glynnis Tidball have both contributed their time.

Lastly, I would like to acknowledge the ongoing support from my husband, Danny Ressl; my mother, Gulshan Valiani; my father, Jafferali Valiani; and my brother, Jahangir Valiani. Their encouragement and love throughout my educational experience was, and still is, greatly appreciated.
1.1 Room Simulation and Auralization

Humans spend a great deal of time listening, learning and speaking in rooms. Rooms differ in geometry and acoustical properties. Some rooms provide better acoustical environments for speech, or other sounds such as music, than other rooms because of these differences. To design better rooms for acoustical purposes, acousticians and architects traditionally predict, using physical or mathematical models, the sound field of the room before it is built. More recently the subjective evaluation of the simulated room sound has become possible by the use of acoustical virtual reality - auralization.

Auralization refers to the process of simulating the sound field in a listening environment, such as a room, and then replaying the simulation to listeners. If auralization is performed correctly, then the signals received at the listeners' eardrums from the simulated room - and the resulting acoustical experience - should be identical to those in the real room. To understand the auralization process, the propagation of sound between a source and a listener in real and simulated rooms must be considered. Figure 1.1 gives an overview of this process.

In the real room, the sound radiated by a source interacts with the complex acoustical response of the room, as well as with the complex auditory system of the listener. The signals that result after these interactions are known as the eardrum signals. In auralization, the source sound signals are digitally recorded and modified by the predicted room response and by an ear response, known more technically as the Head-Related Transfer Function.
Chapter 1. Introduction

Real Environment

Sound Source

Sound Altered by Room's Acousical Response

Sound Altered by the Ears of the Listener

Eardrum Signals

Auralization

Digital Sound Source

Sound Altered by Predicted Room Response

Sound Altered by the Ear Response of the Listener

Replay Compensation Filter

Sound Altered by the Headphones and Played to the Listener

Auralized Eardrum Signals

Figure 1.1: Schematic Describing the Propagation of Sound from a Source to a Listener's Eardrums in Real and Simulated Rooms

(HRTF). The aim is to simulate the complex interactions between the sound signal, the sound field, and the auditory system. To playback the signal to the actual listener, a replay transducer - typically headphones - is used. The headphone’s imperfect response and the passage of sound through the second listener’s ears - that is, the ear canal of the actual listener - lead to distortion of the signal. Thus, the headphone and ear responses must be compensated using an inverse filter during simulation. The final signals represent the "auralized" eardrum signals.

Auralization systems are becoming useful tools for investigating the perceptual performance of the human auditory system. Thus, they are a natural choice for use in hearing research. In the work reported in this thesis, auralization was used in a study aimed at
Chapter 1. Introduction

examining the difficulties in hearing and understanding speech faced by the elderly. These difficulties faced by the elderly are especially apparent in noisy rooms with highly acoustically-reflective surfaces, leading to increased echoes or reverberation. Auralization provides a means of creating a laboratory-controlled sound environment, such as a noisy, reverberant room, for investigating the psychoacoustical capabilities of the elderly. However, before performing psychoacoustical investigations, the auralization technique must be adapted to the purpose and tested extensively to maximize simulation accuracy.

Other important applications of auralization are for concert-hall research and for noise-control in factories. A concert hall or factory is simulated before its actual construction to determine whether its acoustics are appropriate for its type of application - for music, in the case of concert-hall design, or for the detection of speech and warning signals, in the case of a factory. The aim is to allow human subjects to listen to the auralized signals within these virtual environments and make a subjective evaluation of their acoustics. Unfortunately, the auralization process is not perfect. Inaccuracies can be introduced at various stages of the process. It is crucial to understand the perceptual consequences of the inaccuracies.

Testing the accuracy of auralization systems can be done physically - by comparing sound signals in virtual and real rooms - or by subject-based listening tests. Specifically, listener performance during the simulation of an environment is compared to listener performance in the real environment. An important limitation to this approach is that the auralization system, once tested, can be used with some confidence only in procedures or environments similar to those simulated during the validation procedure. For example, if the auralization system showed reasonable accuracy for localization tests in small rooms, then this same accuracy could not be presumed for speech-in-noise tests in large rooms. Once the system’s accuracy is known, it can be used for specific tasks, in specific simulated environments, with a great deal of reliability.

There are a number of auralization systems available today. They range from higher-priced, high-precision systems to low-cost systems. Besides the price/performance gap, the systems vary with their accuracy - for example, in calculating the room response. The room
response can be predicted using many different approaches, which vary in assumptions, accuracies and calculation times.

1.2 Architectural Acoustics and Sound-Field Prediction

A key aspect of auralization is accurate sound-field prediction in rooms. A room is an environment that is enclosed on all sides by partially sound-reflecting walls. The sound-absorptive properties of the surfaces are described by their absorption coefficient - the proportion of sound energy incident on the surface that is absorbed. The energy that is not absorbed is reflected. Two types of reflections occur in rooms: specular reflection and diffuse reflection. In the case of specular reflection, sound that strikes a plane surface reflects from that surface at an angle equal to the angle of incidence. This type of reflection is expected for smooth, hard surfaces. Any acoustical or physical irregularities on the reflecting wall - that is, bumps or grooves that are equal to or smaller than the wavelength of sound - lead to diffuse reflections. In the case of diffuse reflection, a sound ray that strikes a surface at any angle reflects from that surface in all directions. Sound may also reflect from a surface in a partly diffuse manner (see Figure 1.2).

The type of reflections in a room is one room-acoustical parameter that influences room sound-fields and should be taken into account in a comprehensive prediction model. Other factors that influence the sound field are room geometry, the acoustical properties of the room's surfaces, the contents of the room (fittings), the sound-source characteristics, source and receiver positions, and air absorption.
Chapter 1. Introduction

To account for all of these room-acoustical parameters, as well as the infinite number of sound reflections occurring in a room, using a room-prediction model requires long calculation times using the fastest computers. For real-time applications such as auralization, perfect modeling of rooms is, therefore, not possible. Instead, acousticians develop approximate methods to calculate the sound field in rooms. Stated in a more technical manner, the task is to predict the room response resulting at a receiver position from the sound radiated by a sound source, as quantified by the room impulse response (RIR).

The RIR describes the variation of pressure with time at a receiver position that results from an impulsive sound radiated from a source in a room (see Figure 1.3). The first pressure peak of the RIR normally corresponds to the sound that propagates directly to the receiver and does not interact with any surfaces, also known as the direct sound. After the arrival of the direct sound, several signals, altered by a small number of reflections from the different walls, arrive at the receiver - these are called low-order reflections. Later, a multitude of signals arrives at the receiver from various directions as a result of the original signal reflecting from the different surfaces of the room many times - this is referred to as "reverberation".

The RIR can be predicted using many different approaches. Most prediction approaches make the assumption of geometrical acoustics, whereby sound propagates as rays, as opposed to waves. This assumption is expected to be accurate for predicting the RIR at high frequencies. Some approaches, such as the method of images and ray tracing discussed
in more detail in Chapter 2, normally account for specular reflection in a room and ignore diffuse reflection by walls. Another approach, called diffuse-field theory, predicts the RIR assuming a diffuse sound field. A diffuse sound field is such that:

- the reverberant sound waves are incident from all directions with the same intensity at any position in a room;
- the reverberant sound field is the same at every position in the room.

Furthermore, a diffuse sound field can only exist in an empty room that is quasi-cubic in shape with uniformly-distributed surface absorption, or which has uniform absorption and walls that are diffusely reflecting [1]. The diffuse sound field, and the sound field predicted by diffuse-field theory, are associated with an exponential decay of sound. Unfortunately, most rooms do not satisfy the restrictive requirements associated with a diffuse sound field. Therefore, RIR prediction using diffuse-field theory leads to inaccurate results in most rooms.

Two other techniques which have been used to predict the RIR assuming diffuse reflections are radiosity and a modified ray-tracing approach. These techniques yield fairly accurate results in predicting the physical response of real rooms, but they require long computation times to achieve this accuracy.
Chapter 1. Introduction

The portion of the RIR consisting of reverberation - also known as the reverberant tail - is typically characterized by the presence of diffuse sound energy. The energy in this portion can be predicted accurately using algorithms that assume diffuse reflection. There is a need for accurate and practical sound-prediction models, accounting for the effects of diffusion, that can provide objective and subjective information similar to the information obtained in a real room. Developing such a sound-prediction model could lead to better room designs and a better diagnostic tool for hearing research.

1.3 Development of Accurate Sound-Field Prediction Models

Within the context of this research, auralization has two applications:

1. To examine subjects’ ability to localize speech sounds in simulated rooms and other environments, as it pertains to hearing research.
2. To evaluate the sound fields of rooms using a more accurate subjective approach - thereby allowing better room design for particular applications;

Given these applications, the objective of this thesis research was to develop accurate sound-field prediction models for use in auralization systems. The prediction of the diffuse reverberant decay in rooms was a focus of the development of the sound-field prediction model. The accuracy of these models was evaluated subjectively and physically, using subject-based localization tests, by comparing the predicted response of the virtual room with that of the real room.

1.4 Thesis Organization

This thesis is divided into two main sections. The first section focuses on the development of a room sound-field prediction model for use in localization tests using a
Chapter 1. Introduction

commercial auralization system. The second section presents radiosity as an improved sound-field prediction algorithm to predict the room response.

Within Section 1, Chapter 2 provides a detailed literature review of human localization in real and virtual environments. Chapter 2 also discusses the different components of an auralization system, as well as how these components affect simulation accuracy. Chapter 3 describes the commercial auralization system, with emphasis on its hardware and software features and limitations. To reduce software limitations in the existing auralization system, modifications to the system were proposed. Chapter 3 also discusses the different experimental conditions for the localization tests. Chapter 4 deals with the acquisition of data in a real test room, which was subsequently used to simulate the test room by programming the data into the hardware. This is followed by the experimental details and the results of the localization tests, in Chapter 5. As a method of validating the sound-field simulations, the localization-performance results in the virtual room were compared to those of a real test room. A summary of the results obtained using the auralization system is also given in Chapter 5. These results lead to the consideration of a different, combined prediction approach - method of images and radiosity - as a means of more accurately predicting the sound-field in a room. This chapter marks the completion of Section 1.

Within Section 2, Chapter 6 introduces radiosity. The requirements for an algorithm to predict the sound field in a room using radiosity are outlined. Chapter 6 also focuses on digital-signal-processing issues associated with the prediction algorithm - including sampling frequency, accuracy of the approach and filtering. Following discussion of these issues is a description of the operation of the radiosity program. Chapter 7 presents the experimental procedure used in the tests done to validate the program. Program validation, followed by a discussion of the results, is presented in Chapter 8; the responses of several real and scale-model rooms were measured and compared with prediction. Conclusions for the work conducted in Sections 1 and 2 are contained in Chapter 9.
SECTION I:
ROOM SOUND-FIELD PREDICTION AND AURALIZATION FOR HEARING RESEARCH

- CHAPTER 2 -

TECHNICAL BACKGROUND

It is important to investigate how the auditory system localizes sound in an enclosed environment because the reproduction of these localization cues is an important function of an auralization system. Since the focus of this section of this thesis concentrates on using auralization in localization tests to understand the hearing difficulties in the elderly, an examination of human localization in real and virtual environments must be examined.

Various approaches have been used to investigate the mechanisms underlying auditory sound localization by humans. The diversity of these approaches reflects the complexity of the problem. In the next sections, recent studies on auditory localization in real and virtual environments will be reviewed.

2.1 Localization in Real Environments

During the past several decades, a large number of studies have been carried out to investigate the mechanisms involved in auditory localization. Blauert reviewed these studies [2]. Many involved measuring a human subject's spatial impressions in two planes - azimuthal (or horizontal) and median - as illustrated in Figure 2.1. The mechanisms for source localization in these two planes were found to be very different. Within the work reported in this thesis, horizontal-plane localization tests were performed, thus constituting a need to examine studies on auditory localization in the horizontal plane only.
The psychoacoustical perceptions that a human experiences during a localization task in the horizontal plane are dependent on the pressure signals received at the eardrum positions [3][4]. For instance, for a sound source in close proximity to the right ear, the stimulus arrives at the right ear sooner than at the left ear; this normally yields the perception that the sound source is on the right-hand side of the subject. The difference in arrival time of sound at the two ears is called the **interaural time difference** (ITD). ITDs lead to phase differences between the two ear signals. Similarly, the signal at the right ear is slightly more intense than the signal received at the left ear (according to the inverse-square law relationship in physics). This yields an **interaural intensity difference** (IID) at the two ears. The intensity difference caused by the inverse-square law, however, causes very small IIDs. In fact, IIDs are mostly due to head-shadowing effects, which are prevalent for stimulus wavelengths close to or less than the size of the head. Along with IIDs, spectral differences between the two ears are also introduced by the head, as well as by the torso and pinnae.

Current views suggest that the ITDs, IIDs and spectral cues contain the necessary information that aid the auditory system in the task of localization, particularly in the horizontal plane. The next section examines these cues in both anechoic - containing only the direct-sound contribution - and reverberant environments.
2.1.1 Free-Field Auditory Localization

To completely understand a human’s ability to localize in the case of uni-directional sound incidence, many investigators began by analyzing localization perception in an anechoic environment, also known as a free field. Stevens and Newman [5] measured a listener’s ability to localize sound in a free field using pure-tone and noise stimuli in the horizontal plane. They found that a listener’s ability to localize pure tones in the frequency range below 1 kHz and above 4 kHz was better than between 2 kHz and 4 kHz. This gave support to the theory that localization was governed by two mechanisms: ITD and IID cues. Below 2 kHz ITD cues are most important, while above 4 kHz IIDs are most useful. They concluded that the inability to localize tones well in the mid-frequency range was due to the lack of effectiveness of the IID and ITD cues around 3 kHz.

Another finding of this study was the phenomenon of front-back reversals. A front-back reversal describes the situation where a sound arriving from in front is perceived as arriving from the rear, or vice versa. Stevens and Newman determined that the ability of a listener to localize tones directly in front of or behind the listener was equivalent to random guessing for frequencies less than 3 kHz, and slightly higher for higher frequencies. Mills [6] found that the IIDs and the ITDs cues were not adequate in explaining the mechanisms involved in localizing and discriminating sounds from in front or behind, because such sounds yield identical IIDs and ITDs. He argued that the spectral cues supply the auditory system with information on the direction of sound for front/back localization, because the head, torso, and pinnae alter the spectral content of a signal arriving from the front or rear for increasing frequency. Lastly, Stevens and Newman concluded that listeners localize noise more readily than pure tones. They attributed this result to the fact that noise contains both low frequencies and high frequencies, providing concurrent ITD and IID cues.

From Mills and Stevens and Newman, it can be concluded that the auditory system can localize broad-band signals, such as noise or speech, in the azimuthal plane reasonably well in quiet, anechoic conditions. However, the auditory system can also localize speech with excellent performance under noisy and time-varying conditions. Good and Gilkey [7]
examined this phenomenon by testing the sound-localization ability of human observers in the free field, using a variable signal-to-noise ratio ranging from +14 to -13 dB. The noise was always located directly in front of the subjects. As the signal-to-noise ratio was lowered, the accuracy of localization judgments decreased monotonically - the front/back judgments were very negatively influenced by the noise. Despite the monotonic decrease, however, localization in the horizontal plane remained fairly accurate even at low signal-to-noise ratios of approximately -10 dB.

An experiment performed in a nearly anechoic environment, whose results pertain to auditory perception in rooms, was performed by Wallach et al. [8]. Their study found that the direction of an auditory event is determined by the signal that arrives first at the two ears. During the study, subjects initially listened to identical signals played in a room through two loudspeakers at equal distances from the subject and with an angular separation of 60°. With this configuration, the subjects believed that the sound signal was located at the midpoint between the two loudspeakers. Then, a delay of 0 - 35 ms to one of the speakers was introduced. The result was a shift in localization perception towards the loudspeaker without the delayed signal. This phenomenon was named the "precedence effect". Wallach determined that the precedence effect is mainly a function of arrival time, and not of the level of the signal. Furthermore, the precedence effect is best demonstrated with transient sound signals, such as speech or clicks.

The precedence effect partially explains our ability to localize sounds in reverberant environments if one represents the loudspeaker with its signal delayed as a single reflection from a surface in a room. Then, from the precedence effect, this single reflection does not alter the perceived direction of the sound source, represented by the loudspeaker with the non-delayed signal. According to Blauert (1983), the precedence effect is described as an inhibitory process where "components of the ear input signals that originate with the reflection are not fully evaluated; ... their evaluation is totally or partially suppressed." [2]. Rooms, however, contain more than just one or a few reflections, implying that the
precedence effect does not provide a complete understanding of our ability to localize sounds in reverberant environments, such as rooms.

2.1.2 Reverberant-Field Auditory Localization

Testing listeners' localization abilities in the free field can lead to useful knowledge about the human auditory system. However, the ability to localize sound in the free field is rarely exercised in everyday experience. Listeners more commonly localize sounds in enclosed environments containing many reflections, such as in rooms.

As mentioned in the previous section, the precedence effect explains part of the mechanism behind human localization in rooms. Based on this information, investigators began to test the significance of the precedence effect in rooms. In particular, Hartmann [9] published his work on sound localization in the horizontal plane in a variable-acoustics room, using a variety of stimuli. The results of his tests showed that earlier lateral reflections - that is, first-order reflections from the ceiling and floor that have the same azimuth as the direct sound - reinforce the listener's sense of location, while early reflections from the side walls result in confusion in locating the sound source. Hartmann suggested that the very early reflections are likely to be confused with the direct sound and are likely to lessen the precedence effect in rooms.

Rakerd and Hartmann [10] investigated the results from the Hartmann study [9] by evaluating localization in rooms with only one reflecting surface, using both slow-onset and impulsive, 500-Hz sinusoidal tones. Their results indicated that the precedence effect is more observable for localizing impulsive stimuli than with slow-onset stimuli, yet the precedence effect does not completely exclude the effects of early reflections on localization perception. Consequently, any model of localization in a room with many early reflections requires a theory more complex than the simple precedence effect, whose result was obtained in a free field.

Other results obtained from Hartmann's study [9] were that a listener's ability to locate steady-state sounds improved monotonically with increasing spectral density. This
agrees with the results of Stevens and Newman for a free field. Furthermore, the localization of steady-state noise degrades as the reverberation time of the room increases, where the reverberation time is defined as the time at which sound energy in a room from a source which ceases to radiate has decayed to -60 dB. The former result was attributed to the increase of ITD and IID cues presented to the subject, while the latter result is characteristic of the increased masking caused by the reflections in the room for larger reverberation times.

In summary, the ability to localize sounds in an environment is dependent on three acoustic cues: interaural time differences (ITDs), interaural intensity differences (IIDs), and spectral cues introduced by the torso, head, and pinnae. Localization of a sound in the horizontal plane, excluding the front or back, is primarily based on the analysis of interaural time and intensity differences. However, distinguishing the direction of a sound arriving from the front or behind are based on an analysis of the spectral details of the stimulus. In addition, speech is localized better than tones because speech contains both low and high frequencies, which provides sufficient IIDs and ITDs cues. Lastly, the precedence effect, which describes the apparent suppression of the early reflections arriving from the sound source to a listener's ears, plays a role in understanding localization in rooms. However, it does not completely explain auditory localization in rooms.

With this general understanding of the mechanisms involved in localizing sound sources in the horizontal plane in real environments, let us now review what is known about sound localization in virtual environments.

### 2.2 Localization in Virtual Environments

Many studies have investigated simulation accuracy by comparing the results of localization using virtual and real test signals, as described below. These studies generally showed a good agreement between the results for the simulated and real environments, though many of these studies assumed free-field conditions. Very few studies have
investigated simulation accuracy using reverberant-field conditions because of the complexities and difficulties associated with modeling the room's sound field.

Acoustical engineers are striving to achieve better simulations of rooms [11][12], by accurately predicting the RIR. To predict the infinite number of reflections in a room with perfect accuracy is still beyond the capacity of the average computer. Normally, the RIR is predicted up to a certain number of reflections with good precision. Kleiner et al. [13] noted that "at the present it is not known to which extent it is possible to truncate the impulse response."

The sound-prediction model used to construct the RIR can degrade the precision of an auralization system. Other auralization components - the head-related impulse response, replay transducer, and sound source - also lead to errors. The next sections describe the parameters involved in an auralization system, their associated sources of errors, and the effects of these errors on localization ability. The RIR will be discussed last, due to its complexity and importance to the research reported in this thesis.

2.2.1 Head-Related Impulse Response

The head-related impulse response, whose Fourier transform is commonly referred to as the head-related transfer function (HRTF), represents - for a source radiating in the free field - the filtering function of the external ear, the shadowing effects of the head and the reflection effects of the torso. The HRTFs for many individuals have been measured and studied by several investigators [3][14][15][16]. In these studies, HRTFs are customarily measured at or near the eardrum position of a listener with a probe microphone, using an impulsive signal radiating from different source positions around the listener's head in a free field. These HRTFs provide the frequency information (or spectral cues), for a given source position, received at the eardrum. This frequency information is used to locate sounds in an environment, particularly for sources in the median plane. Wightman and Kistler's HRTF
data for three individuals “SOS”, “SJX” and “SOU” are available publicly, and were utilized in the research presented in this thesis.

There are two obstacles in auralization pertaining to HRTFs. The first obstacle deals with using a custom-made HRTF (individualized HRTF) versus using another person’s HRTF (non-individualized HRTF) during auralization. The second obstacle pertains to the inaccuracies of the HRTF at higher frequencies, associated with HRTF measurement.

In relation to the former issue, Wightman and Kistler’s data [17] suggest that listeners’ localization errors are lower when the virtual signals are synthesized using their own HRTFs. Wenzel et al. [16] also investigated the issue of non-individualized HRTFs. Listeners were exposed to virtual sources synthesized from the HRTFs of a good localizer; that is, the HRTFs of a listener who exhibited above-average localization ability in free-field conditions. Their study showed that localization of both free-field and virtual sources was accurate for 75% of the subjects. From this, they concluded that most listeners can obtain useful directional information for azimuthal angle using the non-individualized HRTF of a good listener. Wightman and Tucker [18] have further investigated the effects of using non-individualized HRTFs. They found that the fidelity of virtual signals depended on the similarity between the pinnae used to create the given HRTF, and the listener’s pinnae. For similar pinnae size and shape, localization performance in the real-life situation and the virtual situation were similar. According to Wenzel et al., and Wightman and Tucker’s results, non-individualized HRTFs can give accurate results, for most listeners; inaccuracies are the result of individual differences of the listeners.

The second issue with respect to HRTFs - that of the high-frequency inaccuracy - results from problems with the HRTF measurement in the ear canal. Bronkhorst [19] found that localization of sound stimuli in the frequency range of 7 - 15 kHz was poorer in the virtual than in the real anechoic environment. He postulated that this result followed from an incorrect simulation of high-frequency spectral cues in the HRTFs. He attributed this to distortion in the probe-microphone measurements in the ear canal resulting from the probe-
microphone's finite size and its possible placement at the nodes or antinodes of the ear canal at high frequencies.

The inaccuracies of the HRTFs at high frequencies also result in a distortion of the spectral cues provided by the HRTFs, which could lead to front-back reversals. A study by Begault and Wenzel [20], using non-individualized HRTFs, recorded a mean value of 29% for the percentage of reversed judgments. Another result of their study was the tendency for subjects to perceive sound in the horizontal plane as being elevated.

From the localization studies investigating the effects of HRTFs, it can be concluded that the localization of virtual stimuli synthesized with HRTFs is associated with more front-back confusions, more azimuthal errors and a feeling of source elevation, compared with localization in the real environment. Surprisingly, when using signals synthesized with HRTFs, localization accuracy tends to be better at the rear than at the front [17][20] implying that localization performance is also affected by the azimuth of the source.

2.2.2 Replay Transducer

Most auralization systems replay signals using headphones, though loudspeakers can be used [21]. The replay transducer adds an additional complication to the simulation, namely distortion. Since the eardrum signals created using auralization are designed to be identical to the eardrum signals resulting in the real environment, the replay part of the auralization system must have a flat response; that is, it must not distort the signal. Unfortunately, the auralized signals suffer distortion in amplitude and phase on passing through the headphones. In addition, the HRTFs already contain spectral information about the auditory system. Therefore, replaying signals filtered with HRTFs to a listener will result in a signal affected by two ear canals, one included in the HRTFs and the other associated with the listener's ear. Because this replay and ear distortion can lead to incorrect subjective impressions, it must be corrected for. This replay compensation is normally achieved by the introduction of a compensation filter in the replay network. Ideally, this filter is designed
Chapter 2. Technical Background

such that its transfer function is the inverse of the combined transfer function of the headphones and listener's ear. In practice, the compensation filter only accounts for the magnitude, and not the phase, of the transfer function, since phase differences do not lead to differences in perceptual experience for speech signals [22]. Møller et al. [23] investigated the transfer function of 14 different sets of headphones, to evaluate the headphones as a means for the reproduction of binaural signals. They concluded that none of the headphones were adequate without compensation, since each headphone frequency response contained unique resonances at high frequencies.

The perceptual effects of using headphones for replay were investigated by Begault and Wenzel [19] in a horizontal-plane localization experiment using speech stimuli filtered with non-individualized HRTFs. 15% - 46% of the stimuli were heard inside the head. Furthermore, several subjects showed biases in the azimuthal plane towards the side positions (vertical-lateral plane). These two phenomena have been attributed to the use of headphones during a simulation without accurate replay compensation [24].

2.2.3 Sound Source

To simulate a given source signal radiating acoustically in a given environment, the sound source is altered by the other auralization parameters, such as the HRTFs and RIR. It is preferable that the source signals, to be input to the simulation process, not contain reverberation. In other words, these signals should be recorded in an anechoic environment. Quite often, however, this is not the case with available sound material. Lehnert and Blauert [25] investigated the effects of a sound recording, containing reverberation, on binaural simulation. They stated that there should be no audible differences in the perceived reverberation of the resulting signal as long as the reverberation time of the recording room is less than about 30% of the reverberation time of the simulated room.

Furthermore, Lehnert and Blauert argued that the effects of the recording environment on the structure of the early reflections can be significant if the volumes of the simulated
room and of the signal recording room are very different. For instance, consider two shoe-box-shaped rooms, where the simulated environment has a volume 64 times greater than the smaller recording environment, and the reverberation time of the recording environment is less than 30% of the reverberation time of the large room. In the interval between the arrival of the direct sound and the first reflection in the large room, the convolution of the impulse responses of the two rooms would yield a pattern of early reflections similar to that in the small recording room. Since early reflections may strongly influence auditory parameters such as room spaciousness, one cannot expect the auditory impression to be unaffected by the reflection pattern in the recording room.

In summary, for accurate simulation, the sound-source stimuli must be recorded in a room that satisfies the following two criteria:

1. The recording room's volume should be similar to that of the simulated environment;
2. The recording room's reverberation time should be much less than 30% of the reverberation time of the simulated environment.

2.2.4 Room Impulse Response (RIR)

From an architectural perspective, the most interesting auralization parameter is the room impulse response (RIR), because it describes the way that sound propagates in the room. Historically, the RIR was originally predicted using a scaled-down physical model of the room [21]. Binaural listening in these models is possible using dummy-head recordings [26]. Due to the immense recent progress in computer technology, sound-field prediction by computer has become a convenient alternative to physical scale modeling. Several prediction algorithms have been developed based on geometrical acoustics, whereby sound is propagated as rays rather than waves. The disadvantage of these models is that scattering and diffraction effects, both of which are wave-propagation effects, are ignored.
In geometrical-acoustic prediction algorithms, sound may reflect from a surface either specularly or diffusely. Two different approaches are used for sound-field prediction assuming specular reflection from surfaces: the method-of-images and ray-tracing. Both have been applied to the acoustics of rooms [12][27].

Using the method-of-images technique, a sound source, located next to an unbounded rigid wall, results in a sound-field at a receiver point next to the wall that is the sum of the sound fields generated by the sound source and the image source at this receiver position. This image-source position is determined by mirroring the original source in the plane of the reflecting wall. For environments with multiple surfaces, as in the case of rooms, multiple reflections will occur and, consequently, there will be multiple image sources.

For rectangular rooms, the implementation of this method is straightforward [28]. However, for arbitrarily shaped rooms, the implementation becomes complex, because visibility tests for obstructions must be performed on the image sources [29]. Another disadvantage to the method-of-image approach is that the number of images grows exponentially with the number of reflections in the RIR. As an example, Vorländer illustrated the time required to compute the RIR for an environment with 30 walls, a volume of 15 000m$^3$ and a desired impulse-response length of 400 ms; the computation time was estimated to be 10,000 years on any IBM-AT compatible computer. For these reasons, the method-of-image approach can only be used for very short impulse responses and for simple rectangular rooms for which the visibility test is not required.

In the conventional ray-tracing approach, sound is radiated from a source as dimensionless rays [30] or particles [27], pyramid-shaped beams [31], or conically-shaped beams [11]. These rays, traveling with a certain energy at the velocity of sound, encounter the surfaces of the environment and reflect at an angle equal to the incident angle. Energy is absorbed by the surface as wall absorption. When these rays strike the receiver, information about the rays' energies and arrival times is recorded as a contribution to the impulse response at the receiver location. After tracing all rays, the complete impulse response at the receiver location is obtained.
Chapter 2. Technical Background

One main advantage of ray tracing over the method of images is that ray tracing can, in fact, account for both specular and diffuse reflections in a room. Furthermore, the computation time increases proportionally with the length of the impulse response, and not exponentially as in the method of images. A disadvantage of ray tracing is that temporal resolution is limited because a receiver of non-zero size must be used; which prevents the accurate modeling of strong or isolated sound components caused by specular reflections from smooth walls.

To compensate for the advantages and disadvantages of the method of images and ray tracing, researchers have developed combined method-of-image and ray-tracing approaches [27][32].

In the late part of the RIR, often called the reverberant tail, the energy is mainly contributed by diffuse sound reflections. Thus, it is reasonable to consider sound-prediction approaches based on diffuse reflection, or a combination of diffuse and specular reflections, to generate the reverberant tail.

There are a number of different approaches in modeling the diffuse reflections in a room in the context of sound-field prediction by computer. Kuttruff (1971), as well as Hodgson [33], used ray-tracing techniques that take diffuse reflections into account. In the EARS Auralization Software [12], a statistical and exponential reverberant tail, deduced from the given or measured reverberation times, is appended to the RIR contribution of the calculated specular reflections.

Lewers [31] presented a beam-tracing method, whereby beams with triangular cross-sectional shape are used to approximate spherical radiation from the source. One component of the beam is allowed to reflect from the surfaces specularly; the other component of the beam comprises the diffuse fraction. The diffuse fraction was propagated in the room using a radiative-exchange process, known as radiosity. By determining “form factors” between pairs of surfaces, the distribution of path lengths is determined. Then, by tracing the rays from the receiver to all surfaces, incident-angle and arrival-time distributions are obtained. Subsequently, sound is propagated from surface to surface and to the receiver. The method
Chapter 2. Technical Background

borrows from the computer-graphics field, where radiosity has been used in visualization to create very realistic visual images [34].

Given a method of predicting the RIR accounting for diffuse reflections, the perceptual effects of the RIR can be evaluated. Begault [35] performed such an evaluation in which inexperienced subjects gave localization judgments for headphone-delivered speech stimuli processed by individualized HRTFs with and without synthetic "spatial" reverberation added to the stimuli. In this study, the reverberant portion of the stimulus was modeled as two separate distributions of exponentially decaying noise, one for each ear. According to his results, the spatial reverberation minimized within-head localization, where 25% of the anechoic stimuli, compared to 3% of the reverberant stimuli, were perceived as localized within the head. The reverberation also increased the magnitude of the azimuth localization errors, with mean errors of 12° and 23° for the anechoic and reverberant speech conditions, respectively. Begault concluded that the increase in azimuth error was an indication of the lack of early reflections in the reverberant stimuli, whereby the early reflections aid to reinforce the sound-source direction as described by the precedence effect. Lastly, the spatial reverberation resulted in an overall percentage of reversed judgments of 33%, equivalent to the percentage of reversed judgments with the anechoic speech.

A recent study by Møller et al. [36] involved recording the full RIR and the HRTF of a listener, and then replaying this signal, convolved with speech, to the same listener. It produced a baseline measurement, due to its nearly flawless simulation, for a localization task in a room for any simulation system. Møller et al. found that the localization performance with the individualized HRTFs were equivalent to that in the real-life situation. Therefore, it can be hypothesized that absolute simulation accuracy can be achieved. However, it is still unclear how precise the auralization parameters - in particular the RIR - have to be to achieve accurate simulations before testing elderly people with hearing difficulties using auralization with some validity. Many studies have investigated the effects of using non-individualized, HRTF-filtered stimuli played through headphones, but few have investigated the effects of different RIRs on localization judgment. From the studies evaluating localization
Chapter 2. Technical Background

performance with different RIRs including reverberation, it was speculated that the early reflections play an important role in the human's ability to localize sound. It was also shown that the addition of reverberation leads to less in-head localization during headphone presentation, compared with non-reverberant stimuli.

2.3 Detailed Research Objectives

The motivation for the first phase of this research was to answer the following question: how accurate is a simulation of a room, as evaluated from localization performance in the azimuthal plane, using a predicted RIR including early reflections plus a simplified reverberant tail? This question will be answered by comparing the localization performance in a real room with that in a simulated room.

Using a commercial auralization system, the objective of the Phase 1 was to evaluate the effects of different RIRs on subjects' localization ability. This involved the following steps:

1. Choosing and characterizing a real test room;
2. Modeling the room using different RIRs as follows:
   - A RIR containing only the direct-sound contribution;
   - A RIR containing the direct-sound contributions, along with the six first-order reflections;
   - A RIR containing the direct sound, and an approximate exponential reverberant-decay tail to account for the diffuse reflections in a room;
   - A RIR containing the direct sound, the six first-order reflections and the approximate exponential reverberant-decay tail.

   The accuracy of the auralized signal was determined during localization experiments. The localization performance for the simulations was also compared with localization
performance in the real room, to validate the different sound-field simulations. These localization tests were performed by Pichora-Fuller et al. [37]; their results will be presented in Chapter 5 of this thesis.

The next chapter describes the commercial auralization system in more detail, with its emphasis on the hardware and software features. Modifications made to the system, to compensate for software limitations, are described. Detailed descriptions of the different simulation conditions for the localization tests are also discussed.
AURALIZATION USING THE TUCKER-DAVIS SYSTEM

Hardware equipped with digital signal processors (DSPs) is necessary in creating auralized environments. These devices perform the convolution between one time-response and another; for instance, between the RIR and the speech signal. Typically, the more DSPs are available, the more powerful the system is.

A commercial auralization hardware and software system from Tucker-Davis Technologies (TDT) was used to simulate a real room, to auralize localization-test sounds in the resulting virtual room, and to evaluate the localization performance of listeners in the horizontal plane. This system contained a hardware component, external to a 100486 personal computer (PC), that contained twenty-eight digital-signal-processors (DSPs) capable of convolving time responses. The system also contained several Windows-based software programs, which were installed on the PC. The PC was connected to the DSP card through a high-speed optical data cable for efficient data exchange.

3.1 Hardware and Software Features/Limitations of the Auralization System

The main hardware module containing the DSP board for the TDT auralization system is the PowerDAC (model PD1). As well as the DSP board, the PD1 module contained four 16-bit analog-digital converters (ADC) and digital-analog converters (DAC). The associated hardware components connected to the PD1 are the low-pass anti-aliasing filter (model FT5) and a stereo-headphone/speaker module. A schematic of the hardware system is shown in Figure 3.1.
To create auralized signals, a signal recorded in a free-field (an anechoic signal) is input to the ADC. To avoid distortion and clipping, the input voltage level of the signal, when applied to the ADC, should not exceed 10 V peak-to-peak (p-p). For an adequate signal-to-noise ratio, the input signal’s amplitude was always kept greater than 5 V p-p, making the input signal span most of the dynamic range of the ADC. Once input into the ADC, the input signal is sampled every 20 μs (corresponding to 50 kHz), converting the signal from analog to digital format for further manipulation by the DSPs.

After the anechoic signal is converted into digital format, it is introduced into one or more of the DSPs. The DSP convolves the input signal with a finite-impulse-response (FIR) filter. Filters must be designed to attenuate, to avoid overflow errors in the DSP. The maximum number of taps allowed by a single DSP is 220; anything greater leads to the system malfunctions. The filters used in this work were 128-tap, linear-phase, FIR filters windowed using the Hanning window.

After the necessary DSP operations are performed, the resulting signal is converted back into analog form by the DAC, which has the same specifications as the ADC in terms of sampling frequency and maximum input voltage.

To remove the effects of aliasing that occur during the conversion from the digital to the analog domain, the input signal is passed through a low-pass anti-aliasing filter, with a cut-off frequency of 10 kHz. Lastly, the resulting signal is passed through a headphone module, and output to a pair of headphones.
Chapter 3. Auralization using the Tucker-Davis System

Once the hardware system is properly connected externally, the hardware components located in the PD1 - the DSPs, the ADCs, and the DACs - are internally ("virtually") connected using the interactive software programs of the system. The software package comprises three main programs: FIR, AutoRoute, and SoundStage. The FIR (Finite Impulse Response) program allows the user to create programmable filters for the DSPs. It does so using a graphical user interface with which the user specifies decibel levels for various frequencies, and the number of taps desired for the filter. Once the filter specifications are defined, the filter is created and its associated filter coefficients are calculated. Filter coefficients are then saved to a file readable by the DSP.

The other two software programs, AutoRoute and SoundStage, are responsible for creating simulations. AutoRoute allows the user to create an electronic-circuit schematic for a particular convolution set up and then creates the connections with the actual electronic components of the PD1 module. SoundStage allows a six-sided room containing a listener and up to four omni-directional sound sources to be defined and visualized. SoundStage always requires an AutoRoute file to determine the hardware schematic for a particular series of convolutions in a virtual room.

As provided by TDT, these simulation programs can predict the direct sound and the six first-order reflections for a rectangular-shaped room with constant absorption at all frequencies. To more accurately predict the sound-field in a room, the TDT system was enhanced by adding a reverberant-decay tail to account for higher-order reflections. Furthermore, wall filters with variable absorption for each octave band were implemented to replace the constant-absorption wall filters. The parameters and implementation of the reverberant-decay tail and filters are discussed in subsequent chapters. Lastly, an inaccuracy in the SoundStage program, associated with the way it assumes sound energy decays with distance from a source, was corrected.
3.2 Localization Conditions

Once the modifications were made to the auralization system, several localization test conditions were simulated in the virtual room. These conditions included varying configurations of sound-field simulations, the choice of non-individualized HRTF used, and the possibility of providing feedback to the test subject. These different conditions are discussed in detail below.

3.2.1 Sound-Field Simulation

Four sound-field conditions were simulated in the virtual room, as follows:

- anechoic (direct sound only);
- direct sound and first-order reflections;
- direct sound and a reverberant tail to account for higher-order reflection;
- direct sound, first-order reflections and the reverberant tail.

The direct-sound localization test was performed to obtain results for a virtual anechoic environment. In addition, the results for this condition were compared to results in the literature for both real and virtual environments in order to evaluate simulation accuracy. The free-field results represent the simplest virtual configuration in which localization tests were conducted.

Figure 3.2 illustrates the hardware components necessary in simulating anechoic conditions. The circuit in Figure 3.2 accomplishes the following:
Chapter 3. Auralization using the Tucker-Davis System

1. It converts an analog input signal into digital form;
2. It sends the digital signal to a digital signal processor (DSP) for convolution with the inbound data stream (using DSP[1]). This converts the mono signal into a stereo signal;
3. The resulting convolved signals are further convolved with the headphone-ear compensation impulse response denoted as the "File ..." (using DSP[0]);
4. The signal is converted into two analog signals, via the D/A converters (DACs).

The inbound data stream IB[0] represents a computer calculation of the variation of pressure with the time of sound arriving at a listener. This calculation is performed in SoundStage, where information about the source and receiver position is specified.

The headphone-ear compensation impulse response was created using the TDT FIR software program. This impulse response neutralizes any distortion caused by the headphones and the listener's ear, as described in Chapter 2. A one-channel compensation network was constructed using a 128-tap FIR filter. This filter was programmed into the DSP and was subsequently fed into both channels of the headphones. Equalization was carried out over the frequency range 0 - 11 kHz. The design of the headphone-ear compensation impulse response is discussed in Chapter 4.

The condition involving direct sound and first-order reflections was the next level of simulation complexity. In performing localization tests for this condition, the effects of the early reflections from the different surfaces on localization were investigated. Figure 3.3

Figure 3.2: AutoRoute Diagram for the "Direct Sound Only" Condition
illustrates the hardware configuration necessary for simulating the direct sound plus first-order-reflection condition. The circuit in Figure 3.3 accomplishes the following:

1. It converts an analog input signal into digital form;
2. It sends the digital signal to an eight-tap delay-processing unit that replicates the input digital signal seven times with different delays. The left-side of the delay processor contains the delay information, which is either a constant or is dynamically calculated using inbound data streams; the right-side contains output ports;
3. The output of the delay-processing unit becomes the input to several DSPs which perform convolutions with either the inbound data stream (using DSP[1]) or with a file (using DSP[8 - 13]). The file represents the reflective characteristics of the individual surfaces of the test room;

Figure 3.3: AutoRoute Diagram for the "Direct Sound Plus First-Order Reflections" Condition
4. The signals convolved with the wall impulse responses are further convolved with inbound data streams;
5. The signals from the DSPs are summed and kept in two temporary storage registers (IREG[0 - 1]), representing the left- and right-ear signals;
6. The resulting left and right signals are convolved with the headphone-ear compensation impulse response (using DSP[0]) and converted into analog signals, via the DACs.

The signal output from DSP[1] constitutes the direct sound, while the signal outputs from DSP[2 - 7]) represent the six first-order reflections from the four walls, the ceiling and the floor. The association of a wall with a particular DSP is done using the SoundStage program.

The files containing different wall impulse responses were created using the FIR program. The design of the three filters representing the three different types of surface (ceiling, floor and walls) is discussed in Chapter 4.

The addition of the reverberant-decay tail is the final level of simulation complexity. Initially, the reverberant-decay tail was appended to only the direct sound - forming the third sound-field condition. Here, the reverberant tail can be considered to account for both the low- and high-order reflections. Under this condition, the effects of adding reverberation were investigated.

The final sound-field condition combines the direct sound, the six first-order reflections, and the reverberant-decay tail. With this setting, the effects of the direct sound, the first-order reflections and the higher-order reflections are investigated simultaneously. In addition, localization performance from this condition was compared to that in the real room, since this configuration represents the most accurate simulation of the sound-field in a real room realized using the TDT auralization system.

In creating the final two conditions, the AutoRoute file shown in Figure 3.4 was used to simulate the reverberant tail. The resulting signal was then attached appropriately, as discussed in Chapter 4, to the direct-sound signal (from Figure 3.2) and the combined direct sound/first-order signal (from Figure 3.3). The reverberant-decay tail was created in an
approximate fashion using a series of impulses to simulate an exponential decay with the appropriate decay rate. It is the recursive loop in the AutoRoute diagram that attempts to replicate the higher-order reflections approximately as an exponentially decaying reverberant tail. The circuit in Figure 3.4 accomplishes the following:

1. It converts an analog input signal into digital form, in addition to attenuating the signal by a factor of 0.25;
2. It sends the digital signal to an eight-tap delay-processing unit, which replicates the input digital signal two times with different constant delays (0 ms and 35 ms);
3. The output from the 0 ms delay tap continues to DSP[0] containing a convolving filter representing the diffuse-field HRTF, which was generated from the RMS average of the HRTF magnitudes measured at positions around the head in order to emulate the condition of sound intensity arriving at the eardrum of a listener diffusely - that is, equally from all directions;
4. The second of the two outputs from the delay processor is attenuated by 0.55, delayed by 35 ms and added to the incoming digital signal. The choice of delay (35 ms) and attenuation constant (0.55) are discussed in Chapter 4. The summed signal becomes the input to the delay processor;
5. Steps 3 and 4 are repeated until the output from Step 3 has approximately zero amplitude;
6. The final convolution from the output of Step 3. is performed with the headphone-ear compensation impulse response. Finally the left and right channels are converted from digital to analog form.
The overall AutoRoute circuit design for each sound-field condition was saved as an *.art file. The corresponding binary file, which was read by SoundStage in order to perform the convolutions in a room for given source and receiver positions, was saved as a *.rs file. During each sound-field simulation, the sounds were to be filtered by a set of non-individualized HRTFs, described in the next section.

3.2.2 Non-Individualized HRTFs

The TDT system was equipped with three different HRTF files measured from subjects in the Waisman Centre in Wisconsin [14]. The HRTF, in the form of a filter, accounts for the effect of the reflections from the pinnae (outer ear) and shoulder, as well as the shadowing effect of the head. The HRTFs from Wightman and Kistler’s subjects “SOS”, “SJX” and “SOU” were utilized in the conditions involved in the research in this thesis. These files, initially in binary format, were converted to HRTF files readable by the TDT system using a TDT software program known as WKInterp. WKInterp generated three output files for each binary HRTF file. The first two files contain the HRTFs for each ear. The third file contained the diffuse-field HRTF; that is, the HRTF corresponding to sound arriving from all directions equally. Figure 3.5 illustrates the HRTF for the left and right ears.
Chapter 3. Auralization using the Tucker-Davis System

![Figure 3.5: Magnitude Response of HRTFs for a Source at 0°: (a) Right Ear of SJX, (b) Left Ear of SJX](image)

Figure 3.5: Magnitude Response of HRTFs for a Source at 0°: (a) Right Ear of SJX, (b) Left Ear of SJX

of one individual for an azimuthal angle of 0°. From this figure, it can be seen that the signal that reaches the left and right eardrum have similar, but not equal, spectra.

An important issue regarding HRTFs is the individual variability of these filters; that is, every individual’s ear physiology - and therefore the HRTF - is different. In this project, it was of interest to study the effects of the non-individualized HRTFs on the subjects’ localization performance and compare the findings to the literature. It was also of interest to determine which of the three HRTFs gave the best localization performance within our subject pool.
Chapter 3. Auralization using the Tucker-Davis System

3.2.3 Feedback Condition

The final variable in the localization tests was feedback. Feedback describes the procedure whereby subjects were informed of the correct response immediately after answering, regardless of whether subjects answered correctly or incorrectly. For both the real and virtual conditions, the decision was made that subject's localization performance would be investigated with and without feedback. This condition was investigated in order to monitor whether the subjects were learning any spatial cues.

For all these conditions, the hardware and software input data were chosen to simulate a real test room. The RIR of the existing test room was measured and the input data necessary to simulate the room were determined from the RIR. In addition, and as mentioned in the previous sections, filters were designed to obtain the desired auralized result. The input data, as well as the filter specifications, are detailed in the next chapter.
As discussed in the previous chapter, several conditions were simulated using the TDT auralization system in order to evaluate the localization performance of subjects. To simulate these conditions, parameters characterizing a real room were determined and programmed into the auralization system, forming a virtual-room model. Other data required for creating the auralized eardrum signals, the HRTFs and the compensation filter for the replay transducer (headphones) and listener's ear, were also determined. The acquisition of these data is described below.

4.1 Measurement of the Room Impulse Response

In the work reported in this thesis, the test room was a variable-acoustics chamber - room 369C of the Library Processing Centre at the University of British Columbia. The dimensions of the room are 5.4m x 3.9m x 2.7m high. It features a floor of vinyl tile on concrete, four walls of drywall on 100 mm studs, and a suspended acoustic-tile ceiling. During the localization experiments in the real room, the room contained eight loudspeakers, a chair, a stool, and a chin rest. This room, with its reverberant soundfield, was chosen because it is representative of a room where elderly people have difficulties in hearing and understanding speech.

The impulse response shown in Figure 4.1 was measured in this test room to obtain the reverberation times and information about the absorption coefficients for the different surfaces. The surface-absorption coefficients are required to simulate the conditions
4.1.1. Room Surface-Absorption Coefficients

The absorption coefficients of the different surfaces of the room were estimated using an empirical method. The procedure for this method was as follows:

1. The reverberation time ($T_{60}$) was calculated from the measured RIR of the test room using diffuse-field theory. The RIR was obtained using the MLSSA [Maximum Length Sequence System's Analyzer] system. This system uses a maximum-length sequence (MLS) as an input signal and calculates the RIR using a Hadamard Transform technique [38]. A 32 kHz sampling rate was used and 64k-point RIRs were obtained. A MLS
signal was played into the test room and the RIR was recorded. A special function of the MLSSA system calculated the $T_{60}$ values from the first 10 dB of the sound decay; the resulting quantity is known as the early-decay time;

2. The average absorption coefficient of the room surfaces, $\alpha$, was calculated from the $T_{60}$ values using the Eyring formula where

$$T_{60} = 0.16 \frac{V}{A_{Eyring}}, \text{ with } \frac{A_{Eyring}}{4} = \frac{4}{V} - S \ln (1 - \alpha)$$

in which $V$ is the volume of the room in m$^3$, $S$ is the room surface area in m$^2$, and $m$ is the energy air-absorption exponent in m$^{-1}$;

3. The absorption coefficients of the individual surfaces, $\alpha_i$'s for $i = 1, 6$ were estimated from $\alpha$, the surface areas, and physical considerations.

Given the rigid structure of the floor and walls, as well as their composition, the floor and walls would be expected to have very low absorption. The ceiling absorption coefficients should be higher because of the porous material that is used to make acoustical tiling. Tables of average surface absorptions in octave bands listed in Engineering Principles of Acoustics [39] provided the reference values for the different surface materials used to estimate the $\alpha_i$'s. The absorption coefficients for the empty test room are shown in Table 4.1. These absorption coefficients were programmed into the TDT system using the FIR software program. Three 128-tap FIR filters were created from the results shown in Table 4.1. The magnitude response, $|H|$, of these filters was calculated in octave bands using the equation:

$$|H| = 10 \log(1 - \alpha_i)$$

The filters for the three different surfaces are shown in Figure 4.2.

<table>
<thead>
<tr>
<th>Octave Band</th>
<th>125 Hz</th>
<th>250 Hz</th>
<th>500 Hz</th>
<th>1000 Hz</th>
<th>2000 Hz</th>
<th>4000 Hz</th>
<th>8000 Hz</th>
</tr>
</thead>
<tbody>
<tr>
<td>Floor</td>
<td>0.02</td>
<td>0.02</td>
<td>0.01</td>
<td>0.03</td>
<td>0.03</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>Ceiling</td>
<td>0.35</td>
<td>0.22</td>
<td>0.23</td>
<td>0.30</td>
<td>0.35</td>
<td>0.50</td>
<td>0.55</td>
</tr>
<tr>
<td>Wall</td>
<td>0.09</td>
<td>0.06</td>
<td>0.04</td>
<td>0.05</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
</tr>
</tbody>
</table>
Figure 4.2: a.) Ceiling, b.) Floor and c.) Wall Filters

Chapter 4. Auralization System Input Data
4.1.2 Room Reverberation Times

With the room furnished with eight loudspeakers, one stool, one chair, one chin-rest, and one typical individual - representing the acoustical conditions during the localization tests - the $T_{60}$ values of the room were once again obtained (see Table 4.2).

As mentioned in the previous chapter, a reverberant-decay tail was implemented using a delay-attenuator feedback loop to replicate the higher-order reflections (see Figure 3.4). With this implementation, the $T_{60}$, governed by the decay parameter, is constant at all frequencies. Accordingly, the target $T_{60}$ for the simulation in octave bands was chosen to be 0.6 s, despite the fact that the $T_{60}$ varied with octave band in the test room - as shown in Table 4.2. The rationale for choosing 0.6 s followed from a review of the $T_{60}$ values in Table 4.2. For high frequencies, which provide speech-intelligibility cues, the $T_{60}$ values for the room were of the order of 0.5 s. For low frequencies, which provide the power of the speech signal, the $T_{60}$ values were approximately 0.85 s. The audiologists involved in the work in this thesis suggested that a $T_{60}$ of 0.6 s was a suitable compromise, providing both power and speech-intelligibility cues.

To obtain the target $T_{60}$ in all octave bands while achieving a reasonable sound quality, the impulse delay (35 ms) and decay constant (0.55), as shown in Figure 3.4, were varied. For time delays longer than 35 ms, individual echoes were heard in the speech signal. For time delays shorter than 35 ms, the speech signal contained unwanted high frequencies, resulting in a “metallic” sound quality. By listening to the sound quality with the different values of time delay and decay constant, and comparing it to that in the test room, the time delay of 35 ms and the decay constant of 0.55 were chosen. Using these values, the impulse

| Table 4.2: Reverberation Times of the Furnished Test Room in Octave Bands |
|-----------------|--------|--------|--------|--------|--------|--------|--------|
| Octave Band     | 125 Hz | 250 Hz | 500 Hz | 1000 Hz| 2000 Hz| 4000 Hz| 8000 Hz|
| $T_{60}$ (s)    | 0.72   | 0.94   | 0.85   | 0.56   | 0.55   | 0.48   | 0.49   |
response of the simulated room, and the corresponding $T_{60}$ values, as measured using the MLSSA system, are shown in Figure 4.3 and Table 4.3, respectively.

Referring to the values in Table 4.3, the $T_{60}$ values for most of the octave bands are approximately 0.6 s. Deviations from this $T_{60}$ value, particularly at low frequencies, probably result from inaccuracies in measuring the RIR due to windowing effects.

In choosing a time delay of 35 ms and a decay constant of 0.55, the auralized sound appeared more reverberant than in the real test room. There are three possible reasons for this:

Table 4.3: MLSSA Calculation of Reverberation Times for each Octave Band from the Impulse Response of Figure 4.3.

<table>
<thead>
<tr>
<th>Octave Band</th>
<th>125 Hz</th>
<th>250 Hz</th>
<th>500 Hz</th>
<th>1000 Hz</th>
<th>2000 Hz</th>
<th>4000 Hz</th>
<th>8000 Hz</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{60}$ (s)</td>
<td>0.50</td>
<td>0.56</td>
<td>0.58</td>
<td>0.61</td>
<td>0.62</td>
<td>0.62</td>
<td>0.60</td>
</tr>
</tbody>
</table>
Chapter 4. Auralization System Input Data

1. The speech material was recorded in a sound booth. Thus, the recordings contain some reverberation from the sound booth; these sounds become re-reverberated with the reverberant-tail set up;

2. The SoundStage program simulates omni-directional sound sources; the sources in the real environment are directional, resulting in more sound reflecting from surfaces in the simulated room;

3. The $T_{60}$'s of the test room vary with frequency, whereas the model for the virtual room has a frequency-invariant reverberation time of 0.6 s.

Despite the more reverberant nature of the virtual test-room sound, the values for time delay and decay constant were kept as 35 ms and 0.55, respectively, in order to fulfill the $T_{60}$ requirements.

4.2 HRTFs

As described in the preceding chapter, HRTFs were provided by the University of Wisconsin [14][17]. These HRTFs were measured within the ear canal, at 1-2 mm from the eardrum of the subject, using probe microphones - thereby avoiding standing-wave nulls at high frequencies and capturing all directionally-dependent information. The subjects were seated in an anechoic chamber containing a semicircular arc of loudspeakers located 1.4 m from the centre of the head of the subject. In total, 505 HRTFs, in the form of impulse responses, were measured for each subject. Each impulse response was a 256-point time sequence with a sampling rate of 50 kHz.

Using the coordinate system shown in Figure 2.1, measurements were made at 10 degree intervals of azimuth (ranging from $+180^\circ$ to $-170^\circ$) and elevation (ranging from $+80^\circ$ to $-50^\circ$). The HRTFs used in this thesis were from subjects "SJX", a 1.8 m tall female, "SOU", a 1.7 m tall female and "SOS", a 1.9 m tall male. Figure 4.4 compares the frequency spectra
of the three HRTFs used in the localization experiments in the virtual condition, for an azimuth angle of 90°.

From the curves in Figure 4.4, a considerable amount of variability is apparent at higher frequencies, as described in Wightman and Kistler's paper [14]. As a result of this variability, differences in localization performance among our subjects were expected under virtual conditions.

These HRTFs were programmed into the PD1 module through the program, SoundStage. SoundStage contains a built-in function allowing the user to specify the HRTF desired for the simulation.

![Horizontal Plane HRTFs for an Azimuth Angle of 90°: (a) SJX, (b) SOS, (c) SOU.](image)
4.3 Headphone-Ear Replay-Compensation Filter

Once the signals were filtered by the RIR and HRTF, the signals were replayed through a pair of Sennheiser HD265 headphones. As discussed previously, the headphone and the ear of the listener cause distortion of the eardrum signals; this distortion can lead to incorrect reproduction of subjective impressions. Compensation was achieved by the introduction of a FIR filter in the replay network. The replay headphone-ear compensation filter was measured using the arrangement shown in Figure 4.5.

A Norsonics (NE 830) Real-Time Analyzer was used as both a white-noise source and a FFT (Fast Fourier Transform) spectrum analyzer. White noise was played through the ADC of the TDT PD1 module. Thus, the input signal was digitized and then filtered with an all-pass filter. Then, the signal was converted back into analog form using the DAC output on the TDT PD1 module. The resulting signal was filtered by an anti-aliasing filter (FT5) with a cutoff frequency of 10 kHz, and played into the Sennheiser HD 265 headphones.

The headphones were placed on the ears of a KUNOV dummy head [40][26]. The KUNOV dummy head is a replica of the human head, equipped with microphones at the positions of the eardrums. The terminating impedance of the eardrum is simulated using Zwislocki couplers.

The signals presented to the ears of the Kunov dummy head were received by the Brüel & Kjaer ½” free-field microphone (type 4165) located at the eardrum position of the KUNOV left ear. The microphone signals were transferred to the input of the real-time

![Diagram](image)

Figure 4.5: Set Up for Measuring the Headphone-Ear Replay Compensation Filter
Chapter 4. Auralization System Input Data

The resulting signal was analyzed in FFT mode, and the amplitudes were recorded manually at appropriate frequencies. The transfer function was found to be highly dependent on the position of the headphone on the dummy-head's ear. A method was developed to allow constant and well-defined alignment of the headphone. With this positioning, and correcting for the microphone calibration [free-field microphone in a cavity], the compensation transfer function of the filter was determined (See Figure 4.6).

One observation regarding Figure 4.6 is that frequencies near 2000 Hz are attenuated the most, since the ear canal of the dummy head, as well as of the average human, results in amplification due to a resonance around 2000 Hz. For this and other resonances that occur in the ear canal, the compensation filter must neutralize the amplification effect by attenuating the resonance. It was verified that the final result of a replay headphone, listener's ear, and the compensation filter is a flat response at all frequencies.

One possible inaccuracy associated with the compensation filter results from the fact that the dummy-head ear is only accurate to 7 kHz. At frequencies higher than 7 kHz, the eardrum of the dummy head has a different impedance than that of the average human ear and, thus, presents an inaccurate representation of the ear-canal transfer function at these frequencies.

![Figure 4.6: HD 265 and KUNOV Ear Compensation Filter](image)
Chapter 4. Auralization System Input Data

frequencies. Despite this problem, a one-channel compensation network was constructed using a 128-tap FIR filter, whereby equalization was carried out over the frequency range 0 - 10 kHz. The filter was programmed into the DSPs and applied to both the left and right channels of the headphones.

Once all of the auralization parameters were determined and implemented into the system, localization tests were performed in the real and virtual rooms. A description of the tests performed, as well as their results, is reported in the next chapter.
VALIDATION OF THE ROOM SIMULATION

Once the various simulation conditions were programmed into the TDT auralization system, listeners' abilities to localize speech signals were tested and compared in the real room and in virtual simulations of the same room. Listeners were aged 18-25 with normal hearing and with English as a first language. The next section discusses the experimental arrangement for these localization tests.

5.1 Experimental Arrangement

Localization performance was evaluated for source positions in the horizontal plane. In the real room, eight Tannoy PBM 6.5 II loudspeakers, connected to four dual-channel Alesis RA-100 power amplifiers located in an adjacent room, were arranged in a circle, 45° apart, such that each loudspeaker was 1.52 m from the listener seated in the centre of the room. The heights of the loudspeakers' centres were set to ear level (1.34 m). The real test-room configuration is shown in plan in Figure 5.1.

It was necessary to prevent head movement during the localization tests in the real room, since head movements can produce additional cues that aid in localizing a sound source and since head-movement cues were not present in the auralized signals. An adjustable chair and a chin rest were situated inside the loudspeaker ring to seat the subjects and to minimize head movement during the tests, respectively.

Soundfiles were played in the real room from a 100486 PC and presented through a 16-bit DAC (TDT, model DA3-8). The DAC passed the signal through one of its eight
outputs, through a 10 kHz low-pass anti-aliasing filter (TDT, model FT6), and through an amplifier. The signal was then played to a loudspeaker corresponding to the DAC output.

Before performing any tests, the transfer function of the different loudspeakers and amplifiers were compared, to ensure that the channels gave spectrally equal outputs. This was accomplished by playing a MLS signal through each amplifier/loudspeaker combination. The signals from the loudspeaker were measured with a Brüel & Kjaer ½" microphone (Type 4165) located at the subject position. The frequency responses were analyzed in octave bands and compared for each loudspeaker/amplifier combination, using MLSSA. The frequency spectra recorded for each loudspeaker/amplifier combination, with the amplifier at a constant setting, were equal within ±0.3 dB for each octave band.

Localization tests in the virtual room were conducted in a soundbooth at the School of Audiology and Speech Sciences at the University of British Columbia. The soundfiles were
played from a 100486 PC, presented through a 16-bit DAC (TDT, model DD1), passed through a 10 kHz low-pass anti-aliasing filter (TDT, model FT5) and played through the pair of Sennheiser HD 265 headphones.

5.2 Source Signals and Test Procedure

During the localization tests, twenty soundfiles, each a 4-second segment of 8-talker speech babble, were used. The RMS voltage of all sound files was 2.01 V. In the real room, one soundfile was emitted randomly from one of the eight loudspeakers at a sound-pressure level of 70 dB at the subject position. The subject, with his/her head fixed and facing loudspeaker 1, was asked to identify, using a button box, which loudspeaker the speech was coming from. The subject’s response, as well as the time it took him/her to respond, were sent to a file on a PC, using additional TDT hardware (model SP1). Altogether, the subject was asked to localize the 20 speech-babble files played once from each of the eight loudspeakers; each set of tests thus involved a total of 160 localization judgments per test condition.

Two conditions were used during the real-room localization tests: without feedback and with feedback. Initially, the subjects were asked to identify which loudspeaker emitted the speech without receiving any indication of the correct answer. During the second set of tests, the subject was given the correct response after responding, either correctly or incorrectly, to the speech signal. The tests were conducted in this manner to determine whether the subjects were learning particular cues to assist them in localizing.

For the tests in the virtual environment, the subjects were placed in an audiometric booth with the headphones placed appropriately on their heads. There was no head-fixation device present in the virtual experiments. For each trial, the subjects indicated, using a button box, where the source seemed to be located. Once again, 160 localization judgments were made for each condition. Subjects were tested both without and with feedback during virtual localization.
As previously discussed, four simulation conditions of the real room were simulated using the TDT system: direct sound, direct-sound plus first-order reflections, direct sound plus reverberant tail, and direct sound with first-order reflections and reverberant tail. Before the localization tests were performed, all speech-babble signals used in the real room were convolved with the appropriate RIRs from all loudspeaker positions, one of the three different HRTFs, and the headphone-ear replay compensation filter, using the setup described in Chapter 3. The resulting signals were then recorded and replayed randomly one at a time to the listener during the virtual localization experiment.

The results from the real and virtual rooms are documented in the next section. Due to time restrictions, no complete tests of directional judgment were made with the signals including the reverberant tail. Only the results of informal localization tests are discussed. Despite the lack of virtual results for the reverberant tail, the other localization tests in the virtual room are still useful for comparison with the results of the localization tests that were carried out in the real room.

### 5.3 Results and Discussion of the Localization Tests

The results of the localization experiments in the real room are shown in Figure 5.2. In each diagram, the abscissa represents the loudspeaker presenting the stimulus, and the ordinate represents the loudspeaker perceived to be presenting the stimulus. Shown for each stimulus loudspeaker - by the size of the circle - are the relative magnitudes of the perceived loudspeaker positions averaged over 24 subjects.

In the real room, the perceived and actual directions were in close agreement, yet performance was still less than perfect. Front-back reversals were common; there was a greater tendency to perceive the front-arriving speech as arriving from behind, compared with the reverse situation. Errors in localizing sounds on the left side of the subjects also occurred. Feedback significantly reduced the localization errors from the left-side, as well as front-back reversals.
Some results from the virtual room are presented in Figures 5.3 and 5.4. Figure 5.3 compares the perceived direction as a function of the actual direction for the conditions of direct sound and direct sound plus first-order reflections, using the SOU’s HRTF. The performance in these conditions was considerably poorer than in the real room. Front-back reversals occurred in both sound-field conditions, with fewer back-to-front exchanges occurring with the added first-order reflections. The overall effect of adding the first-order reflections was to decrease localization performance for most positions, particularly at loudspeaker positions 1 (directly in front) and 6 (rear-left quadrant).

Figure 5.4 compares the perceived direction as a function of the actual direction for the direct sound plus first-order-reflection condition using the different HRTFs. For all three HRTF, the localization performance is very poor compared with that in the real room.
Chapter 5. Validation of the Room Simulation

Figure 5.3: Localization Test Results in the Virtual Room Using the HRTF from SOU without Feedback: (a) Direct Sound (b) Direct Sound with First-Order Reflections

Figure 5.4: Localization Test Results in the Virtual Room with Direct Sound and First-Order Reflections without Feedback: (a) Using SOS's HRTF (b) Using SJX's HRTF (c) Using SOU's HRTF

The results using SOS's HRTF display more front-back reversals than the results from SOU's or SJX's HRTF. Also, the localization performance for loudspeaker 4 (rear-right quadrant) using SOS's HRTF was worse than that with SOU's HRTF, whereas the converse was true for loudspeaker 2 (front-right quadrant). From the averages, the best localization
performance was obtained using SJX’s HRTF, yet the performance varies from individual to individual.

According to the informal localization tests involving the reverberant tail, localization errors occurred. The subjects in these informal tests attributed the difficulty in localization to the fact that sound appeared to arrive from all directions simultaneously at a level comparable to the direct-sound contribution. The informal localization-test results using the reverberant tail were inconclusive, but one fact is certain - the implementation of the reverberant tail was crude, even though it was the best that could be done with the system available.

During all of the virtual-room localization tests, the sources were perceived to be located outside the head. Furthermore, when feedback was provided during the virtual tests, localization errors and front-back reversals were reduced significantly [37].

5.4 Consideration of a Different Prediction Algorithm

In general, localization performance in the virtual room was worse than that in the real room. There are two likely explanations for this result: the non-individualized HRTFs used and the different degrees of accuracies with which the RIR was predicted. Studies investigating the effects of non-individualized HRTFs on localization ability have shown that non-individualized HRTFs can give accurate results for most listeners [17][20]. The effects on localization judgment of the different accuracies with which the RIRs are predicted led to inconclusive results. According to Møller et al. [36], when one records the RIR and uses it in simulations with individualized HRTFs, the localization performances in the real and virtual rooms are equal. It can be concluded that a prediction algorithm that can accurately predict the RIR can yield correct localization results. The prediction algorithm for the reverberant tail used in Section 1 consists of a simplified approach that is easy to implement, yet it is a crude approximation to the measured reverberant tail, thereby leading to inaccuracies. To predict the RIR with better accuracy a different algorithm - for example, radiosity - is needed to account for the diffuse-sound energy. This provided the motivation for the second phase of this thesis research.
In Section 1, it was concluded that the technique used for constructing the RIR in the first phase of this work was inadequate for simulating localization tasks using the commercial auralization system. In particular, the implementation of the reverberant tail was crude, despite the fact that it was the best that could be done with the system available. As a result, the development of a better sound-field prediction technique to construct the RIR was proposed. This technique was based on a combined method-of-images and radiosity approach. During the progression of this thesis, it was made aware that the results produced by radiosity could not lead to valid RIRs. Despite this, the model was still developed and room-response measures, such as the octave-band $T_{60}$ values and the steady-state sound-pressure levels - the maximum sound-pressure level achieved by a radiating steady-state source - ($L_p$), were compared between the measured and predicted model's results to validate the predictions.

The concept of the method of images was discussed in Chapter 2.2.4. In this chapter, the background to radiosity, as well as its theoretical development, are presented. Furthermore, design specifications for the new prediction program are discussed. Lastly, a flow chart outlining the different components of the acoustical-radiosity algorithm are presented.
6.1: Literature Review and Critique

Radiosity is a technique used for computing the exchange of radiation between two surfaces. In the 1950s, radiosity was applied to the field of thermodynamics for investigation of radiative heat transfer [41]. Three decades later, radiosity began to gain popularity with the computer-graphics community to model light propagation in rooms [34]. During that same time, Kuttruff derived the radiosity equations for sound propagation in rooms [42]. Even though Kuttruff derived these equations during the 1970s, radiosity has only recently been considered by other acousticians as a means of predicting sound fields in rooms [31].

Conventionally, the method of radiosity assumes that all surfaces in the environment of interest are ideal diffuse reflectors - otherwise known as Lambertian reflectors. These reflectors have the property that the radiation, in the form of heat, light or sound, reflects with equal intensities in all directions, regardless of the angle of the incident radiation. In the lighting realm, newer methods of radiosity have been developed to handle non-diffuse surfaces, such as mirrors, creating more realistic images [43]. These newer methods, however, require the same amount of computational effort and calculation times as ray-tracing techniques. Using radiosity for sound propagation, as compared with light propagation, requires an additional dimension - namely time - since the speed of sound is finite, whereas the speed of light can be presumed to be infinite. Including time in a radiosity algorithm increases calculation times significantly. Thus, predicting sound propagation in rooms using these newer methods of radiosity, which handle specular and diffuse reflections, may lead to considerably longer calculation times.

If one assumes that the surfaces of an environment are ideal diffuse reflectors, radiation-transfer calculations in radiosity are based solely on the geometry of the environment, and not on receiver position. That is, radiosity considers only the interaction of radiation with the surfaces in the environment. Thus, radiation calculations, which may take some time to compute, need only be calculated once. Subsequent to the calculation, a receiver may be positioned at any position in the room and the contributions of radiation from each surface to a receiver can be added in a matter of milliseconds. For this reason, radiosity
is considered as a view-independent process, as opposed to ray tracing in which a new, exhaustive calculation must be made for every new receiver position, making it a view-dependent process.

As stated previously, conventional acoustical radiosity assumes that the surfaces of the room are diffusely reflective. However, room surfaces have been shown experimentally to reflect sound both specularly and diffusely [33][44]. Therefore, radiosity on its own may not be able to predict the sound fields in rooms accurately. Radiosity techniques should, however, be useful in predicting sound fields when combined with a specular approach [31].

Other combined approaches have been used to predict the sound fields in rooms. Hodgson [33] compared predictions using a ray-tracing method that accounted for specular and diffuse reflection with measurements made in a number of rooms, including an empty scale-model room. His results for the scale-model room showed an excellent agreement between prediction and experiment, if the surfaces of his scale-model room were assumed to be 10% - 40% diffusely reflective. As mentioned before, the ray-tracing method is a view-dependent process, suggesting that Hodgson's prediction algorithm must perform a lengthy calculation for each source-receiver position. The radiosity technique, on the contrary, requires less than a minute to obtain the sound-field information for any receiver position, because of its view-independent nature. The radiosity algorithm also provides easy implementation of multiple sources. Thus, a combined specular model and radiosity algorithm should be able to duplicate Hodgson's results with significantly faster calculation times.

Lewers [31] investigated sound-field prediction using a combined ray-tracing and radiosity approach. The ray-tracing component of his algorithm modeled purely specular reflection, whereby each ray was represented by beams with triangular cross sections. The ray-tracing algorithm terminated once the energy radiated by a sound source had decayed by 50 dB. The radiosity algorithm, which Lewers refers to as "radiant exchange", modeled only the diffuse reflections. In combining these two approaches, Lewers allowed the sound signal to be ray traced, as well as diffusely reflected by radiosity, throughout the entire duration of the sound-propagation process. Then, he simply added the energies calculated from both
approaches for all time. Lewers stated that "as time progresses after sound leaves the source, the proportion of energy in the specular model decreases while that in the diffuse model increases." Using this approach, Lewers was faced with long computation times. Calculation times can be reduced significantly if sound is radiated specularly until a set time, after which the sound radiates diffusely. For real-time auralization systems, such an approach, with its low calculation times, are favorable.

6.2 Detailed Research Objectives

The objective of Phase 2 of this thesis was to develop a program to predict the sound field in a room using a combined method-of-images and radiosity approach. The structure of the resulting RIR using the new prediction technique would parallel that of Section 1, with the direct sound, first-order specular reflections calculated using method of images, and the diffuse-reverberant tail calculated using radiosity. This program was then validated by modeling real and scale-model rooms and predicting the sound field at receiver positions within the rooms. The next sections outline the theoretical basis for the radiosity approach, as well as some digital-signal-processing (DSP) issues related to aspects of the new prediction program.

6.3 Theoretical Development

When predicting the sound-field in a room using radiosity, the surfaces of the room are divided into a mesh of elements known as patches. Figure 6.1 illustrates the patches used for the test room in Section 1. From the distribution of patches in the environment, the amount of energy per second leaving a given patch and arriving at a second patch is calculated. This amount of energy, represented as a proportion, is the form factor $F_{ij}$ between patches i and j (see Figure 6.2). Interestingly, this lengthy calculation is independent of whether one is modeling light, heat or sound. Consequently, one can model sound and light using the same form factors.
In order to determine the form factors between two patches \( i \) and \( j \) with areas \( A_i \) and \( A_j \), respectively, consider the geometry between two infinitesimal surfaces with differential area \( dA_i \) and \( dA_j \) (see Figure 6.3). Maintaining the earlier assumption that the sound incident on a patch is diffusely reflected, the radiant energy leaving \( dA_i \) per second which is directly incident on \( dA_j \) is:

\[
\Phi_i \cos \phi_i \; d\omega \; dA_i = \frac{\Phi_i \cos \phi_i \cos \phi_j \; dA_j \; dA_i}{\pi \; r^2}, \quad \text{where} \quad d\omega = \frac{\cos \phi_j \; dA_j}{r^2}.
\]
Figure 6.3: Geometry for Form-Factor Derivation (extracted from Goral et al., 1984 [45])

$\Phi_i$ is the intensity (in Watts/m$^2$) emitted from patch $i$, $r$ is the distance between the centers of the patches, and the angles $\phi_i$ and $\phi_j$ represent the angle between the normal vector of the patch and the vector $r$.

Noting that the total energy per second leaving patch $i$ is $\Phi_i dA_i$, the form factor representing the fraction of the total energy per second flowing from $dA_i$ which is directly incident on $dA_j$ is:

$$F_{ij} = \frac{\Phi_i \cos \phi_i \cos \phi_j dA_j dA_i}{\pi r^2} = \frac{\cos \phi_i \cos \phi_j dA_j}{\pi r^2}$$

Moreover, the form factor between the finite surfaces with areas $A_i$ and $A_j$, which is the sum of the contributions from the infinitesimal areas, is defined by:

$$F_{ij} = \frac{1}{A_i} \int_{A_i} \int_{A_j} \frac{\cos \phi_i \cos \phi_j dA_j dA_i}{\pi r^2}$$

Simple identities that reduce the form-factor calculation time exist [45]:

1. Given two patches with diffusely-distributed reflected and emitted energy, a reciprocity relationship exists whereby

$$A_i F_{ij} = A_j F_{ji}.$$
Thus, given \( F_{ij}, A_i \) and \( A_j, F_{ji} \) can be easily determined;

2. Given a plane or convex surface, \( F_{ii} = 0 \);

3. Given a closed environment of \( N \) patches, all the energy leaving a surface must be accounted for according to the conservation of energy. Therefore, the form factors for each surface must sum to unity.

For an enclosure with \( N \) patches, leading to a form-factor matrix with \( N \times N \) elements, the above simplifications can reduce the form-factor calculation time by approximately one-half.

The form factors are an integral component in solving the radiosity equation. They provide information on the amount of sound, light, or heat energy a given patch receives. Given the form factors for a particular enclosure, one can proceed to evaluate the radiosity equation.

In radiosity, both the energy received by a given patch and the energy reflected from that patch are important. The prime objective in radiosity is to determine the intensity \( B \) leaving a surface; this depends on the energy received and reflected from a patch.

The system equation for radiosity is simply:

\[
B_j = E_j + \rho_j \sum_{i \neq j} B_i F_{ji}
\]  

(6.3)

where \( B_i \) and \( B_j \) represent the intensities, or *radiosities* of patch \( i \) and patch \( j \), respectively, \( E_j \) is the intensity emitted from patch \( j \), and \( \rho_j \) is the reflection coefficient of patch \( j \). The equation states that the power, or the energy per second (\( B_j \)), leaving a unit area is the sum of the power emitted from that patch (\( E_j \)) and the power reflected as a result of the incident power contributed by all other patches (\( \rho_j \sum B_i F_{ji} \)).

To apply Eq. (6.3) to acoustics, two modifications are made. The first involves considering air absorption; the second accounts for the sound-propagation time between two patches. Air absorption describes the phenomenon when the magnitude of a sound wave decreases when propagating from one surface to another through the propagation medium. Air absorption is described by the following formula:
Air Attenuation Factor = $\Phi_{ji} = \exp(-m r_{ji})$ \hspace{1cm} (6.4)

where $m$ is the air-attenuation exponent and $r_{ji}$ is the distance between patches $i$ and $j$. Air absorption in acoustics is significant and should not be ignored. Including air absorption in the radiosity equation yields the following expression:

$$B_j = E_j + \rho \sum_{i \neq j} B_i F_{ji} \Phi_{ji} \hspace{1cm} (6.5)$$

Sound-propagation time differences result from the finite sound speed in air.

Consider sound propagation from patch $i$ to patch $j$ as illustrated in Figure 6.4. As a result of the finite speed of sound, sound energy from patch $i$ propagating towards patch $j$ will move from position $P_0$ (at time 0) to position $P_1$ (at time 1), to position $P_2$ (at time 2) and so on. The energy from patch $i$ (at time 0) will be received at patch $j$ at a time $T_{ij}$, where $T_{ij} = r_{ij} / c$ is the propagation time for a sound wave's energy to propagate from patch $i$ to $j$, with $c$ the speed of sound in air, 343 m/s.

Figure 6.4: Sound Propagation from Patch $i$ to Patch $j$

(Extracted from Shi et al., 1993 [46])
In an enclosed room with \( N \) patches, \( N-1 \) patches contribute energy to any given patch. The \( N-1 \) patches may contribute sound energy to a given patch at different times \( T_{ij} \). It is apparent that at time \( t \), a single patch \( j \) may only receive energy from the \( N-1 \) patches that radiated their energy prior to the propagation time \( (t - T_{ji}) \). As a result, the acoustical-radiosity equation, taking into account both air attenuation and sound-propagation time, becomes:

\[
B_j = E_j + \rho_j \sum_{i \neq j} B_{i(i,T_{ij})} F_{ji}(\tau) \Phi_j \tag{6.6}
\]

### 6.4 Digital-Signal-Processing (DSP) Issues

An algorithm derived from Eq. (6.6) was used to develop the acoustical-radiosity program presented in this thesis. To take advantage of the efficiency of the digital-signal processors (DSPs) found in most auralization systems, the new room-prediction program was designed to output digital signals. The output signal from one run of the program produced a positive-pressure versus time response, also known as the echogram, for any one octave band. The program was run several times to produce the echograms for the various octave bands. Because these echograms were digital, they could be easily manipulated with other sequences, such as speech, using a DSP.

To prepare the program's output digital signal for subsequent auralization, the digital signal's sampling frequency must be chosen appropriately. Once the signal has been sampled, the signal must be filtered correctly to construct the RIR for auralization. This section discusses the choice of sampling frequency, and the filtering technique applied to the echogram outputs. The treatment of wall reflections in the context of filtering is also discussed.
6.4.1 Sampling-Frequency Determination

According to Fourier theory, the time domain and frequency domain are complementary; that is, a change in the time domain causes an effect in the frequency domain, while the same change in the frequency domain leads to an equivalent effect in the time domain. Because of this relationship between the time and frequency domains, the sampling frequency must be chosen to satisfy both frequency and time requirements.

The frequency requirements follow from Nyquist's Theorem, which states that the relationship between the sampling frequency, $\omega_s$, and the maximum frequency accurately transmitted, $\omega_{\text{max}}$, is $\omega_s/2 = \omega_{\text{max}}$. For example, if $\omega_s$ is 8000 Hz, then the maximum frequency whose level can be accurately transmitted is 4000 Hz; the frequency spectrum above 4000 Hz will be a mirror copy of the spectrum in the frequency range 0 - 4000 Hz. Typically, for speech applications, $\omega_s$ is chosen to be 20 kHz, but for the work in this thesis, $\omega_s$ was determined from the highest octave-band frequency used to describe the surface absorption within the program. Surface-absorption coefficients are typically given in octave bands from 125 - 4000 Hz. The highest octave band spans the range 2828 - 5657 Hz; therefore, the maximum frequency that can have a uniquely specified pressure is 5657 Hz. As a result, the minimum sampling frequency, $\min(\omega_s)$, is $11,314 \text{ Hz} = 2(5657) \text{ Hz}$.

Typically, $\omega_s$ is chosen to be slightly higher than the $\min(\omega_s)$.

For a given sampling frequency, the time-domain response contains many sampled points separated by a time interval, $1/\omega_s$. Therefore, $\omega_s = 11,314 \text{ Hz}$ yields a time interval of 88 $\mu$s. Given that a time interval of less than 50 ms between two impulses presented to the ear leads to the perception of one signal, rather than two individual impulses [47], using a spacing of 88 $\mu$s is excessive for playback to a listener's ears. Nonetheless, the echograms should be sampled at a sampling frequency corresponding to 88 $\mu$s, in order to meet the stringent frequency requirements.
In the work reported in this thesis, $\omega_s$ was chosen to be 12 kHz for full-scale room prediction; this is slightly higher than $\min(\omega_s)$, but leads to similar echogram sizes as with $\min(\omega_s) = 11,314$ Hz. To reduce storage requirements during the binaural playback process, a low-pass filter can be applied to the binaural signal to filter out any undesirable frequencies that may result in aliasing, and then decimate the signal to reduce $\omega_s$. For example, in [11] decimation of the binaural impulse response was done to take into account the limited spatial and temporal resolution of the ear, reducing the storage requirements by several hundred kilobytes.

6.4.2 Treatment of Wall Reflections and Filtering

As described previously, the simulation process is repeated in the various octave bands. Each room surface is assigned one absorption coefficient corresponding to the octave-band echogram being predicted. Energy from the source is reduced at each reflection according to the absorption coefficient of the surface. For an impulsive source signal, containing a flat spectrum at all frequencies, the magnitude of the spectrum is attenuated on each reflection by an amount corresponding to the absorption coefficient (see Figure 6.5). To convert these echograms into a RIR, the individual octave-band echograms are filtered with an ideal band-pass filter and combined by superposition [48].

![Diagram of frequency spectrum](image)

Figure 6.5 Frequency Spectrum of a Signal at Different Times
Unfortunately, filtering with an ideal band-pass filter introduces a non-causal signal \[48\] - that is, one which begins prior to \(t = 0\). For the purposes of comparing the predicted and measured RIRs for different rooms, linear-phase octave-band FIR filters were designed which kept the signal causal and maintained time relationships between the different frequencies.

To avoid the use of filters, the wall may also be described by its impulse response. After a sound source radiates an impulsive signal, the energy from the source is convolved with the impulse response of the wall at every reflection \[48\]. Since the program developed in this thesis divides the room surfaces into patches, and since a signal from one patch is incident on nearly every other patch, this approach for wall reflections can lead to excessive calculation times. Therefore, the treatment of wall reflections as filters is most advantageous for our type of simulation.

In most studies \[11\]\[48\], including this thesis, wall filters were represented by their magnitudes or by the average octave-band absorption coefficients - phase was ignored or assumed to be zero. In reality, walls possess both magnitude and phase responses. A recent study \[22\] suggests that, for auralization purposes, our hearing is virtually insensitive to phase relations in sound signals presented in rooms and, therefore, the phase response of a surface may safely be neglected.

### 6.4.3 Discretization of the Surfaces

An assumption of the radiosity approach is that each surface emits energy, in the form of intensity, as a point source. Given that the surface size of, say, a wall or ceiling is probably too big to be modeled as a point source, especially for small surface-receiver distances, the surface is divided into smaller surfaces, known as patches. The size of these patches, however, must satisfy certain criteria in order for them to act as point sources. Following is a discussion of when a finite surface can be modeled as a point source.

In the visual realm, a finite-plane Lambertian emitter can be modeled as a point source only when the distance to the receiving surface is greater than \textit{five times} the maximum
Chapter 6. Technical Background

projected width of the emitter [49]. In the acoustical realm, however, a finite-area plane source can be modeled as a point source when the distance to the receiving surface is greater than $1/\pi$ of the maximum projected width of the emitter [50]. The logic for this argument is as follows: Imagine a planar acoustic source with an area of $a \times b$, where $a > b$, with an observer situated at a distance, $c$, on the vertical axis of symmetry (see Figure 6.6). The square of the sound pressure generated by a point source is given by:

$$p^2 = W \frac{\rho_o c'}{4\pi r^2}$$

where $W$ is the power of the source, $r$ is the source-receiver distance, and $\rho_o c'$ is the characteristic impedance of the medium ($\rho_o c'$ for air is 414 Pa.s/m$^2$ or 414 rayls at standard temperature-pressure). For the finite plane source the sound pressure produced by the surface is

$$p^2 = 2 \int_{x=0}^{a} \int_{y=0}^{b} \frac{W}{a \cdot b} dx dy \frac{\rho_o c'}{4\pi r^2}.$$ 

If we substitute the following variables:

$$x = c \tan \alpha, \quad y = c \tan \beta, \quad r = \frac{c}{\cos \alpha}, \quad R = \frac{r}{\cos \beta}$$

Figure 6.6 Planar Sound Source and Observer (Source-Observer Distance = $c$)
the integral can be written as:

\[ p^2 = \frac{W \rho c'}{\pi bc} \int d\alpha d\beta = \frac{W \rho c'}{\pi bc} \tan^{-1} \left( \frac{a}{2c} \right) \tan^{-1} \left( \frac{b}{2c} \right). \]

From this result, three ranges for which the pressure-squared function is different can be distinguished. The first range describes the case when the observer is close to the source - that is, \( c < b \). The second range is defined by \( a < c < b \). The final range is characterized by \( c > b \).

From investigation of the last range, where \( c > a \) and \( c > b \), the following approximations can be made:

\[ \tan^{-1} \left( \frac{a}{2c} \right) \approx \frac{a}{2c}, \quad \text{and} \quad \tan^{-1} \left( \frac{b}{2c} \right) \approx \frac{b}{2c}. \]

This leads us to a pressure-squared relation of

\[ p^2 = W \frac{\rho c'}{4\pi c^2}, \]

which is equivalent to that for the square of the sound pressure for a point source. The limiting value of \( c \) for which point-source behavior is possible is when:

\[ \tan^{-1} \left( \frac{a}{2c} \right) \approx \frac{a}{2c}, \quad \text{which occurs at} \quad \frac{a}{2c} = \frac{\pi}{2}, \]

\[ \therefore c = \frac{a}{\pi}. \]

As a result, a finite-area plane source can be modeled as a point source when the distance to the receiving surface is greater than the maximum projected width of the emitter divided by \( \pi \).

To confirm this derivation, the number of patches in a room, and thus the patch sizes, were varied to determine whether the simulation results would change for different patch sizes meeting the above criterion. The surfaces of one room were divided into either 96 patches or 150 patches - corresponding to patch areas on the order of 1 m\(^2\) and 0.65 m\(^2\),
respectively. With both sizes of patch sizes, the reverberation time, $T_{60}$, and the steady-state sound-pressure level, $L_p$, were evaluated from the echogram using Eq. (6.7):

$$\text{SP} = 10 \log_{10} \int_0^\infty \frac{\text{Echogram}(\text{amplitudes in pressure}^2)}{(p_0)^2}$$

$$\text{Sound Decay}(T_{60}) = 10 \log_{10} \int_{T_{60}}^\infty \frac{\text{Echogram}(\text{amplitudes in pressure}^2)}{(p_0)^2} - \text{SP} \ (6.7)$$

$$= -60 \text{ dB}$$

where $\text{Echogram}$ represents the amplitudes of the echogram, and $p_0$ is a reference pressure valued at $2e-5 \text{ Pa}$. The latter expression in Eq. (6.7) for “Sound Decay” normally represents a linearly decreasing function with time, especially for diffuse-sound fields. Because of the constant slope of the sound-decay curve, the $T_{60}$ was evaluated by determining the slope of the sound-decay curve between the -10 dB-point and -15 dB point, and then by multiplying the result by 12.

Comparing the calculated reverberation times and the steady-state sound-pressure levels, the different patch sizes predicted the same reverberation times, and similar steady-state sound-pressure levels differing by less than $\pm 1.0 \text{ dB}$. These results confirm the notion that as long as the patch size meets the above criterion, the predicted sound field varied little with further between increases in patch size. A more detailed description of these results is presented in Chapter 8.

### 6.4.4 Echogram Length

For all computer sound-field simulations, the echogram is calculated up until a defined time. This time can be based on the length of the echogram, its energy, the order of reflection - as in the case of the method of images - or a combination of these three factors. As previously mentioned in Chapter 2, there is no definitive criterion presently available to determine when the echogram should be truncated [13]. In this work, the truncation time was derived by the echogram shape. The program developed in this thesis produces echograms
whose shape consists of a direct-sound impulse, several milliseconds of impulsive fluctuations, followed by a smooth monotonic-decay curve. The monotonic decay curves for the rooms investigated in this work led to near-zero amplitudes around 0.3 - 0.5 seconds. Since longer truncation times result in a significant increase in calculation times, the simulation was set to terminate at 0.3. During the progression of this thesis, it was made aware that the echograms containing energy past 0.3 s resulted in slight differences in $L_p$s, but significant differences in $T_{60}$ values. Upon the observation that the monotonic decrease of energy in the echogram can be easily approximated by a correction curve with the same decay rate as the echogram, a correction curve can be appended to the echogram to account for the missing energy past the truncation time, for the purposes of calculating accurate $L_p$ and $T_{60}$ values.

With these design specifications fixed, the new room-prediction program was implemented. The description of the different program components is discussed in the next section.

6.5 MATLAB Program

The program was written using MATLAB v4.2. The program, which typically ran on a SUN Microsystems SPARC 4 Workstation, could be run on any computer system with sufficient memory and disk space. It consisted of two main MATLAB-program modules. A schematic representation of the modules is shown in Figure 6.7.

The first module (designated as "1"), representing the radiosity approach in calculating the RIR, comprises six MATLAB m-files (see Table 6.1). The first, second and third m-files discretizes the surfaces of the room, determines the area of the discretized elements, calculates the distances between elements, and calculates the form factor for each patch, respectively. The fourth m-files radiates the sound power from the source to the patches and then re-radiates the power from each patch to each other patch in a recursive fashion. The fifth m-file sums the energy contribution from each patch at the receiver
Figure 6.7: Program Flow Chart for Predicting the RIR Using Radiosity and Method-of-Images Approaches

position, creating the echogram. Finally, the sixth m-file performs octave-band filtering on the echograms and constructs the RIR for the given source and receiver positions in a room.

The second module, used to construct the RIR using a method-of-images approach, comprises two m-files. The first m-file calculates the echogram, for a given source and receiver position in a room using the method-of-images approach; the second m-file performs octave-band filtering on the echogram response and constructs the RIR for the given source and receiver positions in the room.
There are a number of extra steps required in constructing the RIR using the radiosity technique, as compared to the method-of-images technique. These extra steps result from the discretization of the surfaces of the room and associating each patch with the other patches in the room. To obtain a better understanding of the extra steps involved using the radiosity algorithm, the functions of the different radiosity m-files are briefly discussed in Table 6.1.

Sample input data for the program, used to construct the RIR for a room, is shown in Appendix A.

With the complete program, the sound fields at receiver positions in several real rooms were predicted. A discussion of the rooms and the experimental procedures is discussed in the next chapter.
### Table 6.1: Description of Radiosity M-Files in the First Module of the Program

<table>
<thead>
<tr>
<th>Order</th>
<th>M-File</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>element_size.m</td>
<td>This m-file calculates the vertices for each patch in a room given the dimensions of the room. The number of patches in the room is determined using the criteria, described in 6.4.2, related to the distance between each patch and the largest dimension of each patch.</td>
</tr>
<tr>
<td>2</td>
<td>area.m</td>
<td>Given a set of vertices for a series of patches describing the room environment, this m-file determines the area of each patch and the distance between each pair of patches for the entire room. The output is a nxn matrix, PatchArea_ij, where n represents the number of patch elements in a room.</td>
</tr>
<tr>
<td>3</td>
<td>ff.m</td>
<td>For any radiosity algorithm, the fraction of energy leaving one patch and incident on another patch must be determined. This fraction of energy is known as the Form Factor, F_{ij}, between two patches. This m-file calculates the form factors between each pair of patches and outputs a nxn matrix, FF. The inputs to this file are the vertices of the patches and their areas, PatchArea_{ij}.</td>
</tr>
<tr>
<td>4</td>
<td>fullrad.m</td>
<td>This routine calculates the intensity of sound as a function of time for each patch generated by a radiating point source. This routine uses a two-step process: 1. The point source radiates sound to all patches; 2. Each patch behaves as a secondary source and radiates sound energy to each other patch until a maximum time. The inputs to this file are the form factors, patch areas, source location and powers, and the reflection coefficients of the rooms surfaces. The output is a matrix, B, containing the energy emitted from each patch at discrete times.</td>
</tr>
<tr>
<td>5</td>
<td>t_delay.m</td>
<td>This m-file constructs the echogram at a receiver position using the results from fullrad.m. The $T_{60}$s and $L_p$s are calculated from the echogram. The inputs to this file are the matrix B, and receiver location.</td>
</tr>
<tr>
<td>6</td>
<td>script.m</td>
<td>This routine takes the echograms for each octave band and performs octave-band filtering in the frequency domain. The inverse FFT of the filtered signal is performed to produce the RIR.</td>
</tr>
</tbody>
</table>
This chapter presents the experimental procedures used in the tests done to validate the new prediction model. RIR measurements were performed in four rectangular, empty rooms: two full-scale and two 1:8-scale models. Because scale-model rooms were used, it is of interest to review scale-modelling principles briefly.

7.1 Scale-Modelling Principles

To model the sound field in an enclosure at $1:n$ scale, where $n$ is the scale factor, all dimensions are scaled by $1/n$. All quantities referring to the scale model are denoted with a subscript "SM", such that the corresponding full-scale lengths become $l = nl_{SM}$. Since the velocity of sound in air is the same in both full-scale and scale-model rooms, the relationships between the full-scale and model wavelengths and frequencies are $\lambda = n\lambda_{SM}$ and $f = f_{SM}/n$. The frequencies of the test signals must also be scaled by $n$, so that scale-model wavelength-to-dimension ratios are equal to those at full-scale. The scale-model room used in this research was considered to have a scale factor of $n = 8$. Therefore, with full-scale measurements made in the 125 - 4,000 Hz octave bands, the test measurement range in the scale model corresponded to the 1,000 - 32,000 Hz octave bands. Because of limitations associated with the measurement devices (discussed in the next section), the upper range of the octave-band frequencies for scale-model measurements was in fact limited to 16,000 Hz.

Scaling distances by $1/n$ in the scale model, results in sound-pressure level, $L_p$, and sound-power level, $L_w$, both scaling by $n$. Therefore, a source that produces a certain mean-
square pressure at a distance $r$ in the full-scale room will produce the same mean-square pressure at a distance $r/n$ in the scale model [51].

7.2 Power Calibration and Measurement System

Before performing tests in the various rooms, the sound power of the sound source, in octave bands, was determined using the measurement system illustrated in Figure 7.1. The MLSSA [Maximum Length Sequence System Analyzer] system generated a MLS, which was filtered using a digital equalizer and then amplified by a power amplifier. An omnidirectional loudspeaker converted the amplified signal into sound energy. The sound wave propagated in a free-field environment (anechoic chamber) to the microphone. The microphone converted the incident sound pressure to an electrical signal and transmitted this signal to a measuring amplifier, via a microphone preamplifier. This electrical signal was returned to the MLSSA system after being amplified by the measuring amplifier.

Figure 7.1: Measurement System for Calibrating the Power of the Source
Chapter 7. Experimentation

The instruments used were as follows:

- Portable personal computer, PC III, containing the MLSSA system;
- Digital equalizer, Yamaha DEQ7, with settings appropriate for reproducing a flat output spectrum (see Table 7.1) and a maximum octave-band setting of 16,000 Hz;
- Power Amplifier, QSC Audio USA 370;
- Two different omni-directional loudspeakers used to cover the different frequency ranges:
  - Full scale - dodecahedral loudspeaker array equipped with twelve loudspeakers, Realistic model 40-1284E, with a frequency range of 700 - 20,000 Hz;
  - Scale model - mid-range/tweeter loudspeaker, Realistic model 40-1289A, with a frequency range of 250 - 25,000 Hz, with a cone narrowing to a 3 mm diameter opening to make the source omnidirectional at the model test frequencies.
- Two different microphone/preamplifier/measuring-amplifier systems were used for the different frequency ranges:
  - Full scale - ½" microphone and sound level meter, Rion model NA-29 and NA-29E;
  - Scale model - ¼" microphone, Brüel & Kjaer type 4135, preamplifier, Brüel & Kjaer, and measuring amplifier, Brüel & Kjaer type 2804.

To calibrate the full-scale source power, the sound-pressure level incident on a microphone located 1 m from the source was recorded. The source was rotated to obtain its average sound-power level using the following formula:

\[ L_w = \bar{L}_p + 11 \quad \text{dB} \quad (7.1) \]
Table 7.1: Digital Equalizer Settings in 1/3 Octave Bands for Full-Scale and Model Rooms

<table>
<thead>
<tr>
<th>1/3 Octave Band (Hz)</th>
<th>Full-Scale Equalizer Setting (dB)</th>
<th>Scale-Model Equalizer Setting (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>63</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>80</td>
<td>18</td>
<td>0</td>
</tr>
<tr>
<td>100</td>
<td>7.5</td>
<td>0</td>
</tr>
<tr>
<td>125</td>
<td>6.0</td>
<td>0</td>
</tr>
<tr>
<td>160</td>
<td>7.0</td>
<td>0</td>
</tr>
<tr>
<td>200</td>
<td>2.0</td>
<td>0</td>
</tr>
<tr>
<td>250</td>
<td>-1.0</td>
<td>0</td>
</tr>
<tr>
<td>320</td>
<td>1.2</td>
<td>0</td>
</tr>
<tr>
<td>400</td>
<td>-2.6</td>
<td>0</td>
</tr>
<tr>
<td>500</td>
<td>1.4</td>
<td>0</td>
</tr>
<tr>
<td>630</td>
<td>2.2</td>
<td>1.2</td>
</tr>
<tr>
<td>800</td>
<td>-0.4</td>
<td>4.6</td>
</tr>
<tr>
<td>1000</td>
<td>-1.4</td>
<td>3.0</td>
</tr>
<tr>
<td>1250</td>
<td>-3.4</td>
<td>-0.6</td>
</tr>
<tr>
<td>1600</td>
<td>1.8</td>
<td>1.4</td>
</tr>
<tr>
<td>2000</td>
<td>0.6</td>
<td>1.4</td>
</tr>
<tr>
<td>2500</td>
<td>6.5</td>
<td>-6.0</td>
</tr>
<tr>
<td>3200</td>
<td>2</td>
<td>-0.4</td>
</tr>
<tr>
<td>4000</td>
<td>2.2</td>
<td>1.4</td>
</tr>
<tr>
<td>5000</td>
<td>-5.0</td>
<td>-6.5</td>
</tr>
<tr>
<td>6300</td>
<td>5.8</td>
<td>3.4</td>
</tr>
<tr>
<td>8000</td>
<td>-1.6</td>
<td>-2.0</td>
</tr>
<tr>
<td>10000</td>
<td>4.4</td>
<td>-5.0</td>
</tr>
<tr>
<td>12500</td>
<td>0</td>
<td>-2.0</td>
</tr>
<tr>
<td>16000</td>
<td>0</td>
<td>-16.0</td>
</tr>
</tbody>
</table>

where $L_w$ is the sound-power level in dB, while $\bar{L}_p$ is the average sound-pressure level as measured by the MLSSA system.

Calibration of the scale-model source power was performed using a loudspeaker-microphone distance of 12.5 cm corresponding to the full-scale loudspeaker-microphone distance divided by the scale factor, 8. The source was also rotated to obtain its average
sound-power level using a formula similar to Eq. 7.1 with the added term: 20\log(r), where \( r \) corresponds to the loudspeaker-microphone separation of 12.5 cm.

For each measurement using the MLSSA system, the signal arriving at the microphone was recorded five times, and averaged, using the Go Average MLSSA command. The MLSSA operational parameters for the full-scale rooms were as follows: Acquisition length = 65536 samples (1.819 s); Acquisition sampling rate = 36.04 kHz; Stimulus amplitude = ±0.06152 V; Anti-aliasing filter bandwidth = 10 kHz. The MLSSA operational parameter for the scale-model rooms were: Acquisition length = 65536 samples (1.152 s); Acquisition sampling rate = 75.50 kHz; Stimulus amplitude = ±0.06152 V; Anti-aliasing filter bandwidth = 25 kHz. The sound power, in Watts, for each loudspeaker is shown in Table 7.2.

7.3 Tests Environments

7.3.1 Full-Scale Rooms

The test room, referred to as the Environmental room, used for the localization tests in Section 1 of this thesis, was used to validate the model. Measurement conditions were typically 23 °C and 51 % relative humidity. The air-absorption exponents (Np/m) are listed in Table 7.3. The dimensions of the room, and the absorption coefficients of its surfaces are discussed in Chapter 4. Using the omni-directional loudspeaker, RIR measurements were

<table>
<thead>
<tr>
<th>Octave Band (Hz)</th>
<th>125</th>
<th>250</th>
<th>500</th>
<th>1000</th>
<th>2000</th>
<th>4000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full-Scale Source Power (W)</td>
<td>3.20e-4</td>
<td>6.33e-4</td>
<td>1.35e-3</td>
<td>2.25e-3</td>
<td>4.65e-3</td>
<td>5.71e-3</td>
</tr>
<tr>
<td>Octave Band (Hz)</td>
<td>1000</td>
<td>2000</td>
<td>4000</td>
<td>8000</td>
<td>16000</td>
<td>-</td>
</tr>
<tr>
<td>Model-Source Power (W)</td>
<td>1.28e-3</td>
<td>2.34e-3</td>
<td>1.77e-3</td>
<td>4.41e-3</td>
<td>1.11e-2</td>
<td>-</td>
</tr>
</tbody>
</table>
Table 7.3: Air Absorption Exponents used in Predictions of the Environmental Room

<table>
<thead>
<tr>
<th>octave-band frequencies (Hz)</th>
<th>125</th>
<th>250</th>
<th>500</th>
<th>1000</th>
<th>2000</th>
<th>4000</th>
</tr>
</thead>
<tbody>
<tr>
<td>air absorption exponent (Np/m)</td>
<td>7.75e-5</td>
<td>2.68e-4</td>
<td>7.05e-4</td>
<td>1.32e-3</td>
<td>2.39e-3</td>
<td>5.94e-3</td>
</tr>
</tbody>
</table>

made at three receiver positions in the empty room, as shown in Figure 7.2. Octave-band $T_{60}$ and $L_p$ values were also determined from the RIR for the three receiver positions. $T_{60}$ and $L_p$ values are compared with those predicted in Chapter 8. The RIR was filtered to contain only

Figure 7.2  Source and Receiver Positions in the Environmental Room:
(a) Receiver at (2 m, 2.36 m, 1.29 m), (b) Receiver at (2 m, 2.36 m, 1.5 m), and (c) Receiver at (1 m, 2.36 m, 1.29 m). Source position in all three cases was (2 m, 4.86 m, 1.29 m).
Chapter 7. Experimentation

those frequencies contained in the 125 to 4000 Hz - that is, 88 - 5657 Hz. The resulting signal is referred to as the "wide-band" RIR.

Tests were also done in a large classroom - room 12 of the Hebb building at the University of British Columbia. The dimensions of the room were 13.7m x 7.8m x 2.6m high, as shown in Figure 7.3. Three blackboards, with areas of $7.4m^2$, $7.4m^2$, and $7.2m^2$, were attached to three of the four walls. Curtains, with a width of 2.7 m and a length equivalent to the height of the room, were covering a section of a wall. The walls were made of painted concrete, the floor was linoleum tiles on concrete and the ceiling was acoustical tiles on concrete. The average absorption coefficients, estimated using the process detailed in Chapter 4.1.1, are shown in Figure 7.4. This room was chosen for measurement due to its rectangular geometry and the lack of permanently attached desks and chairs, allowing the room to be completely empty during measurements. Using the omni-directional loudspeaker, RIR measurements were made at three receiver positions in the empty room, as shown in Figure 7.5, to construct the wide-band RIR.

7.3.2 Scale-Model Room

Measurements were also performed in a rectangular, empty 1:8 scale-model workroom. The model was 3.75 m long, 1.875 m wide, and 1.25 m or 0.625 m high, as shown in Figure 7.6. In full-scale dimensions, it was 30 m long, 15 m wide, and 10 m or 5 m high. The floor was made of concrete, representing concrete in the full scale, the ceiling was 1/8” varnished plywood, representing a light-weight steel deck in the full scale, and the walls were made of 3/4” plywood, representing blockwork or brick at full scale. Measurement conditions were typically 19.5 °C with 56 % relative humidity. The corresponding air-absorption exponents (Np/m) for the scale-model rooms are listed in Table 7.5. The average absorption coefficients of the surfaces, estimated from measured reverberation times, are shown in Figure 7.7.
Table 7.4: Air Absorption Exponents used in Predictions for Hebb 12

<table>
<thead>
<tr>
<th>octave-band frequencies (Hz)</th>
<th>125</th>
<th>250</th>
<th>500</th>
<th>1000</th>
<th>2000</th>
<th>4000</th>
</tr>
</thead>
<tbody>
<tr>
<td>air absorption exponent (Np/m)</td>
<td>7.75e-5</td>
<td>2.68e-4</td>
<td>7.05e-4</td>
<td>1.32e-3</td>
<td>2.39e-3</td>
<td>5.94e-3</td>
</tr>
</tbody>
</table>

Figure 7.3: Dimensions of Hebb 12

Figure 7.4: Absorption Coefficients for Hebb 12’s Surfaces
Figure 7.5  Source and Receiver Positions in Hebb 12: (a) Receiver at (5.3 m, 3.9 m, 1.3 m), (b) Receiver at (12 m, 3.9 m, 1.3 m), and (c) Receiver at (11.7 m, 5.8 m, 1.3 m). Source position in all three cases was (1.5 m, 3.9 m, 1.3 m).
Figure 7.6: Dimensions of the Scale-Model Room

Table 7.5: Air Absorption Exponents used in Predictions (Scale-Model Values)

<table>
<thead>
<tr>
<th>Octave-Band Frequencies (Hz)</th>
<th>1000</th>
<th>2000</th>
<th>4000</th>
<th>8000</th>
<th>16000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air Absorption Exponent (Np/m)</td>
<td>1.49e-3</td>
<td>2.57e-3</td>
<td>5.55e-3</td>
<td>1.69e-2</td>
<td>6.01e-2</td>
</tr>
</tbody>
</table>

Figure 7.7: Average Absorption Coefficients of Scale-Model Room’s Surfaces
Chapter 7. Experimentation

Using the omni-directional scale-model source, RIR measurements were made at the same two different receiver positions in both model rooms, as shown in Figure 7.8. The wide-band scale-model RIRs included all frequencies contained in the 1000 Hz to 16 kHz octave band, corresponding to the full-scale frequencies contained in the 125 to 2000 Hz octave bands.

The wide-band RIRs and the associated sound-decay curves are presented in Chapter 8. Comparisons between measured and predicted reverberation times and $L_p$s are also presented in Chapter 8.

Figure 7.8 Source and Receiver Positions in Scale Model Rooms: (a) Receiver at (1.13 m, 0.875 m, 0.10 m), (b) Receiver at (1.9 m, 0.875 m, 0.10 m) Source position in both cases was (0.63 m, 0.875 m, 1.40 m).
This chapter reports the results of predictions by the new room-prediction model, in the form of $T_{60}$ values, $L_p$s, echograms, and RIRs. These results are compared with predictions using alternative approaches. Furthermore, the measured and predicted results are compared, and a discussion of the discrepancies between the measured and predicted results is presented.

8.1 Patch-Size Criterion Validation

In Chapter 6.4.3, a criterion for patch size was presented, which stated that a patch can be modelled as a point source as long as the maximum dimension of the patch is less than $1/\pi$ times the distance from the patch to the receiver. This criterion was validated by comparing the $T_{60}$ values and the $L_p$s for one room discretized into different numbers of patches and, therefore, of patch sizes. The room used for validation purposes was the full-scale environmental room. Each surface dimension was divided either into four or five segments, resulting in 96 or 150 patches, or patch areas on the order of 1 and 0.65$m^2$, respectively. Both patch sizes satisfied the above criterion in this room. The octave-band $T_{60}$ values and the $L_p$s at receiver position 3 (see Figure 7.2) are shown in Figure 8.1. Figures 8.1a and b show, respectively, the octave-band $T_{60}$ values and $L_p$s for the different room-patch sizes. The effect of patch size on the octave-band $T_{60}$ values is small. There are differences in $L_p$ values for different patch sizes, especially at low frequencies; however, for higher frequencies, the differences are very small. These differences in the low frequencies may be a result of the short echogram truncation time preventing accurate representation of the $L_p$ values. To verify this claim, the truncation length for the echogram describing the room
with 96 patches was extended from 0.3 seconds to 0.6 seconds. The resulting $L_p$ values from the longer echogram were similar to those of the room with 150 patches. Therefore, the $L_p$
values have converged for the case where the dimensions of the room were divided into five, corresponding to room containing 150 patches. For fewer patches in the room, longer echograms are required to obtain accurate results for $T_{60}$ and $L_p$ values.

To understand the discrepancy in $L_p$ values between two different patch sizes, consider two surfaces, one divided into two patches, and the other divided into three, as illustrated in Figure 8.2. For both cases, consider rays radiated by the source which are incident on each patch. In the first case, two rays will arrive at the receiver position with certain energies. These energies arrive at two different times, $t_1$ and $t_2$. In the second case, the rays will arrive at the receiver at three different times, resulting in a smearing of energy in time compared with the first case. The energy arriving at the receiver should be similar in both situations if the energy is integrated over time. Because a room with fewer patches contains more energy per time bin, shorter truncation times may result in more energy loss, as compared with the case of a room with more patches whose energy is more evenly distributed. Therefore, increasing the truncation time of the echogram for the room with fewer patches should lead to less differences in $L_p$ values between the two patch sizes, as was seen from Figure 8.1. For the remainder of the tests, the dimension of each surfaces was divided into five segments, giving 150 patches per room.

![Comparison of Sound Propagation Between a Source and a Receiver for Different Patch Sizes.](image)

where $t_1' < t_1$
8.2 Comparison of Radiosity with Diffuse-Field Theory

Because the radiosity model assumes diffuse reflections in rooms, it should predict results similar to those of diffuse-field theory for rooms that satisfy the requirements for a diffuse-sound field (see Chapter 1.2) - that is, a cubic, empty room with uniform surface absorption, whose walls reflect diffusely.

To test this hypothesis, predictions for an empty, cubic room with uniform surface absorption were made using the radiosity model. The room was a hypothetical 3 m x 3m x 3m enclosure with surface absorption coefficient, \( \alpha = 0.6 \), for all surfaces. Comparisons of the predicted \( T_{60} \) and \( L_p \) with those calculated using diffuse-field theory were made. Diffuse-field theory for the \( L_p \) and \( T_{60} \) can be expressed as follows:

\[
T_{60} = 0.16 \frac{V}{A_{Eyring}}, \quad L_p(r) = L_w + 10 \log \left[ Q/4\pi r^2 + 4(1 - \alpha)/A_{Eyring} \right]
\]

with

\[
A_{Eyring} = 4mV - S \ln (1 - \alpha)
\]

where \( r \) is the source-receiver distance, \( L_w \) and \( Q \) are the source sound-power level and directivity factor, \( V \) and \( S \) are the room volume and surface area, \( \alpha \) is the average diffuse-field surface absorption coefficient and \( m \) is the energy air-absorption exponent.

For \( L_w = 93.5 \text{ dB} \), \( Q = 1 \), and \( m = 1.32 \times 10^{-3} \), results for the different source-receiver distances were predicted and calculated (see Table 8.1). The radiosity \( T_{60} \) predictions are accurate within \( \pm 10\% \), while the \( L_p \) values are accurate within \( \pm 0.2 \text{ dB} \). The deviations in the predicted results are within the bounds of “engineering accuracy” [1]. It can be concluded, therefore, that the radiosity model can predict a diffuse field for appropriate rooms.
Table 8.1: Radiosity Prediction versus Diffuse-Field Theory
Results for a Source at [1, 1, 1] = [x, y, z]

<table>
<thead>
<tr>
<th>Receiver Position [x,y,z]</th>
<th>Radiosity $T_{60}$ (s)</th>
<th>Diffuse-Field $T_{60}$ (s)</th>
<th>Radiosity $L_p(r)$ (dB)</th>
<th>Diffuse-Field $L_p(r)$ (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[0.5, 0.5, 0.5]</td>
<td>0.096</td>
<td>0.090</td>
<td>85.1</td>
<td>84.9</td>
</tr>
<tr>
<td>[1.5, 0.5, 1]</td>
<td>0.098</td>
<td>0.090</td>
<td>86.3</td>
<td>86.3</td>
</tr>
<tr>
<td>[2.5, 2.5, 2.5]</td>
<td>0.099</td>
<td>0.090</td>
<td>79.9</td>
<td>79.9</td>
</tr>
<tr>
<td>[0.1, 0.1, 0.1]</td>
<td>0.095</td>
<td>0.090</td>
<td>81.7</td>
<td>81.6</td>
</tr>
<tr>
<td>[2.9, 2.9, 2.9]</td>
<td>0.099</td>
<td>0.090</td>
<td>79.7</td>
<td>79.5</td>
</tr>
</tbody>
</table>

Figure 8.3: Echogram Predicted Using Radiosity for a Receiver at [2.9, 2.9, 2.9]
8.3 Radiosity Validation with Other Prediction Models

Many assumptions and approximations, in terms of surface absorption (see Chapter 6.4.2), were made in constructing the radiosity model of the room. These approximations can lead to inaccuracies in the results. Additional inaccuracies may result from the fact that the real source and the receiver are not dimensionless nor omnidirectional as presumed in the model. Furthermore, the source loudspeaker and microphone have their own transfer-function characteristics - that is, they will distort the measured signal in a room. Even though an attempt was made to compensate for the loudspeaker and microphone octave-band responses, in order to obtain a flat output spectrum, individual peaks and dips in the magnitude spectrum were present; the integration process of the individual frequencies to achieve octave bands, results in a smoothing of discontinuities in the frequency spectrum, but does not remove the discontinuities.

A better approach to validate the radiosity model is to compare the radiosity results with the results of predictions by other prediction models. In this way, similar approximations and assumptions about the room are made with each model, leading to better comparisons. In this research project, four prediction models were used: the radiosity model, developed in this thesis; the method-of-images model, which was later combined with radiosity to constitute the new room-prediction model; a model based on specular ray-tracing techniques; and a model based on diffuse ray-tracing techniques. Using all four models, octave-band $T_{60}$s and $L_p$s were determined from the predicted echograms and compared with results for the environmental room. The echogram for the 1000 Hz octave band, as well as the wide-band RIR, for each method were also evaluated.

The ray-tracing model, RAYCUB, developed by Ondet and Barbry [52] and modified to account for diffuse surface reflections, as described by Kuttruff [53] were used. Further details of this model are found in [33]. The receiver in this model is a cube of non-zero side length, as compared with the radiosity and method of images that use point receivers. Thus, the resulting sound-pressure levels are averages over the cube. When using RAYCUB, the user must choose input parameters such as the receiver size, the number of rays and the
maximum number of ray trajectories to ensure adequate prediction accuracy. In the present work, the receiver size was 0.05m, the number of rays was 1,000,000, and the maximum trajectory was 50.

To achieve the same accuracy as with the ray-tracing model, method-of-images predictions were made up to the 25\textsuperscript{th} image order, corresponding to the maximum trajectory of 50.

The octave-band $T_{60}$ values at all receiver positions in the environmental room, for all prediction approaches, are shown in Figure 8.4. One trend apparent in Figure 8.4 is the similarity between the $T_{60}$ values for radiosity and diffuse ray-tracing. This result is expected since both models assume purely diffuse reflections in rooms. Small differences in $T_{60}$ between radiosity and diffuse ray tracing may have resulted from different method used to approximate the $T_{60}$ in both methods. Furthermore, it is surprising to note that the two models that assume purely specular reflections - that is, method-of-images and specular ray-tracing - lead to very different $T_{60}$ results. In general, the specular ray-tracing model yields $T_{60}$ values that are consistent with the fact that diffuse reflections tend to reduce $T_{60}$ values. The $T_{60}$ values predicted by the method-of-image model, however, appear to be very inaccurate as compared with the results from the specular ray-tracing model. This would suggest that the method of calculating the $T_{60}$ values in the method-of-images model may be in error.

The octave-band $L_p$'s at all receiver positions in the environmental room, for all prediction approaches, are shown in Figure 8.5. The two ray-tracing models, as well as the method-of-images model, are predicting similar results. The radiosity approach, however, underestimates the $L_p$s as compared with the other models. This result may be attributed to echogram length. In all predictions, the sampling rate was set to 12,000 Hz, while the echogram length was variable. Both ray-tracing models and the method-of-images model constructed 5000-point echograms, corresponding to a maximum echogram time of 0.417 s, whereas the radiosity model had a maximum echogram time of 0.3 s. The discrepancy in $L_p$ values between the other models and radiosity suggests that there is energy unaccounted for by radiosity, due to its shorter echogram length.
Figure 8.4: Predicted Reverberation Times from Various Models for the Three Receiver Positions in the Environmental Room: (a) Receiver Position 1, (b) Receiver Position 2, (c) Receiver Position 3; *— radiosity model, x— diffuse ray-tracing model, o— specular ray-tracing model, and +— method-of-images model.

The 1000 Hz echograms for each prediction model, measured at receiver position 3, are presented in Figure 8.6. Generally, the envelopes of the wave forms are similar for the different models. Interestingly, in the case of the radiosity model, sound decays monotonically with increasing time. Furthermore, after a certain time, the energy predicted using the radiosity
The echograms for the ray-tracing models are more sparse than the radiosity echogram. This is a consequence of the limited number of rays used to construct the echogram, as well as the fact that a ray may reflect from many surfaces before arriving at a receiver. Therefore, there are times at which the receiver will not receive any energy, leading to the discrete, decaying, impulse train shown in Figure 8.6 (b) and (c).

The echogram for the method-of-images model, in Figure 8.6 (d) resembles that of the ray-tracing models in the beginning. For increasing time, however, the echogram amplitudes
remain greater than zero until the sound energy has completely decayed. This effect is a result of the image-source density exceeding the sampling rate, such that it appears as if energy is arriving at the receiver continuously. During this decay, the amplitude of the echogram fluctuates significantly, as compared with the echogram predicted by radiosity.

Using these 1000 Hz echograms, as well as those for the other octave bands, the wideband RIRs were constructed, using the method discussed in Chapter 6.4.2. The predicted RIRs for receiver position 3, using the different prediction models, are shown in Figure 8.7. Two effects result from applying octave-band filters to the echogram: the result has positive and negative amplitudes, characteristic of room impulse responses; a shorter signal results, because energy contributions from some frequency bands have been removed by the filtering - this extra energy was contributing to the larger amplitudes and longer lengths seen in the echogram. The latter consequence illustrates an unfavourable attribute of the radiosity prediction model, regarding its use in auralization. To elaborate, the echogram from the radiosity model had a smooth non-fluctuating decay. The frequency response associated with such a smooth time decay contains a large amount of energy at low frequencies. When the echogram is filtered, most of the energy contained in the low frequencies is removed, resulting in a sparse, short signal. The other prediction approaches contain amplitude fluctuations in the echogram. These fluctuations lead to increasing energy at high frequencies. Since most of the energy at high frequencies - between the 2000 and 4000 Hz octave bands - is maintained, these RIRs have much more energy then the RIR predicted from radiosity, leading to longer signals. It is also worth mentioning that measured RIR has a sparse shape similar to those of the different prediction models, excluding radiosity. If the phase of the walls were taken into account in these models, the results would be different. But since phase was ignored, the new room-prediction model proposed here is not suitable for auralization, since the radiosity component of the model leads to inaccurate RIR representations. This new prediction approach can, however, be used to determine $T_{60}$'s and $L_p$'s with similar accuracy to the other prediction methods since these acoustical parameters are obtained from the octave-band echograms.
Figure 8.6: 1000 Hz Echogram Predicted using Various Models: (a) Radiosity, (b) Diffuse Ray-Tracing, (c) Specular Ray Tracing, (d) Method of Images
Figure 8.7: Predicted RIR using Various Models: (a) Radiosity, (b) Diffuse Ray-Tracing, (c) Specular Ray Tracing, (d) Method of Images
8.4 Sound-Field Prediction Model Validation with Measurements.

The RIR's and sound-decay curves for the four real rooms, as measured at receiver position 3 in the full-scale rooms and receiver-position 2 in the scale-model rooms, are shown in Figures 8.8 - 8.11. Even though the objective of this section was to develop an improved sound-field prediction model and compare the RIR from the predictions with those from the receiver, the results of Section 8.3 revealed that wide-band RIR comparison using the new model is inappropriate. Instead, the predicted $T_{60}$ and $L_p$ values using the new model were compared with experiment. Before comparing these physical parameters, it is of interest to consider the echograms predicted by the combined method-of-image and radiosity model. Figures 8.12 - 8.15 show the predicted echogram for the 1000 Hz octave band for each room, at receiver position 3 for the full-scale rooms and at receiver position 2 for the scale-model rooms. From the predicted echograms, the six first-order reflections can be clearly seen, with a reverberant tail, as predicted using radiosity, appended to the final

![Figure 8.8](image)

Figure 8.8: (a) Impulse Response of the Environmental Room, (b) Energy Decay Curve, as Measured at Receiver Position 3
Figure 8.9: (a) Impulse Response for Hebb 12, (b) Energy-Decay Curve for Hebb 12, as Measured at Receiver Position 3

Figure 8.10: (a) Impulse Response the 10 m height Scale Model Room, (b) Energy-Decay Curve for the 10 m height Scale Model Room, as Measured at Receiver Position 2
Figure 8.11: (a) Impulse Response the 5 m height Scale Model Room, (b) Energy-Decay Curve for the 5 m height Scale Model Room, as Measured at Receiver Position 2

Figure 8.12: Echograms for the Environmental Room: (a) Predicted Response for the 1000 Hz Octave Band, (b) Measured Wide-Band Response
Figure 8.13: Echograms for the Hebb 12: (a) Predicted Response for the 1000 Hz Octave Band, (b) Measured Wide-Band Response

first-order reflection. The measured echograms are shown beside the predicted wide-band echograms. In most cases, the direct-sound contribution has similar amplitudes when predicted and measured. The measured echograms also contain very large distinct peaks after the direct-sound contribution; these may be attributed to the transfer function of the loudspeaker source and microphone receiver.

Figure 8.14: Echograms for 10m High Scale-Model Room: (a) Predicted Response for the 8000 Hz Octave Band, (b) Measured Wide-Band Response
Figure 8.15: Echograms for 5m High Scale-Model Room: (a) Predicted Response for the 8000 Hz Octave Band, (b) Measured Wide-Band Response

Despite the obvious differences in the appearance of the echograms, general conclusions cannot be made without comparing the predicted physical parameters of the room, or performing a subjective evaluation of the signal, using auralization techniques. The percentage deviations between the measured and predicted $T_{60}$ values for the different rooms at each receiver position are shown in Figure 8.16. The results show deviations of 0 - 45%. Because the results vary between rooms and frequencies, it is difficult to determine any one reason for the differences. We can only generalize by saying that a number of approximations were made about the room in order to construct these models and it is probably a combination of these approximations that leads to the discrepancy in the results.

The difference in $L_p$'s portrays a different story. The difference in measured and predicted $L_p$ values for the different rooms at each receiver position are shown in Figure 8.17. At most, the differences between the measured and predicted $L_p$'s are 4 dB for the environmental room, and within 10 dB for the remaining rooms. Smaller differences in the $L_p$ for the environmental room, as compared with the other rooms, may be due the combination between the patch size chosen and the echogram's truncation time. In Section 8.1, a discussion of the effects of patch size and
Figure 8.16: Percent Deviation of T₆₀ for all Four Rooms as a Function of Frequency:
(a) Environmental Room, (b) Hebb 12, (c) Scale-Model Room with 10 m Height,
(d) Scale Model Room with 5 m Height; *-- Receiver Position 1, x-- Receiver Position 2,
°-- Receiver Position 3
Chapter 8. Experimental Validation

Figure 8.17: Difference Between Measured and Predicted $L_p$ for all Four Rooms as a Function of Frequency: (a) Environmental Room, (b) Hebb 12, (c) Scale-Model Room with 10 m Height, (d) Scale Model Room with 5 m Height; *-- Receiver Position 1, x-- Receiver Position 2, °-- Receiver Position 3
echogram length was presented. During this discussion, it was stated that accurate $L_p$ values can be obtained for increases in echogram length or for increases in the number of patches in the room. During our predictions, the number of patches in each room was 125, while the echogram truncation time was 0.3 sec. While these parameters yielded reasonable $L_p$ values for the environmental room, they may not be adequate for the remaining rooms. To obtain reasonable accuracy in predicted $L_p$ values for the other rooms, an increase in patch number, an increase in echogram truncation time, or both, may be required.

Both the $T_{60}$ and $L_p$ results from the new room-prediction model exceeded the error allowed for in "engineering accuracy". Despite the noted differences in physical data, it would still be of interest to evaluate these results subjectively by convolving the RIR or echogram from the new room-prediction model with speech and replaying the results to listeners. Unfortunately, because of time restrictions, no subjective listening tests were performed using this room-prediction model.
The objective of this research was to make a contribution to room acoustics with the aim of simulating the sound fields in rooms. The main goal of the research was to develop accurate sound-field prediction models for use in auralization systems. To achieve this objective, two phases of work were undertaken.

In the first phase, outlined in Chapters 2 - 5, acoustical signals were created and presented to listeners using the Tucker-Davis Technologies (TDT) auralization system in order to test sound-localization ability. This involved measuring the room impulse response (RIR) of the environment that was being simulated, using data from the RIR to create the virtual environment, performing localization tests in the real and virtual environments, and comparing the localization performances.

The TDT system was inadequate to use for localization tests without suitable modifications. The work in phase I of this thesis concentrated on modifying the existing system by adding a reverberant tail to account for higher-order reflections, by introducing wall filters that varied with frequency, and by programming a replay-compensation filter to account for distortion arising from the replay-headphones and listener's ear.

In the localization tests with the modified system, front-back reversals were common, with fewer back-to-front exchanges occurring with the direct sound plus first-order reflection condition. Localization error increased for most positions when going from the direct-sound condition to the direct sound plus first-order-reflection condition. In the real room, the localization performance was significantly better compared with the results for the virtual room. Feedback reduced the localization errors in both the real and virtual rooms.

Despite the efforts made to improve the auralization system, the localization-test results in the virtual environment were inconclusive, since inaccuracies attributable to the non-individualized Head-Related Transfer Function (HRTF) and the predicted sound field
both contributed to localization errors in some manner. As a result, further modifications must be made to the system or its parameters. These modifications could include the creation of a facility to measure individualized HRTFs, or the development of a better room sound-field-prediction model, or both. The second of these three options was investigated in the second phase of this thesis.

In the second phase, outlined in chapters 6 - 8, a sound-prediction algorithm was developed, based on acoustical radiosity, and used in the development of an improved, combined, room-prediction model. Because of the innovation of using radiosity in acoustics, the radiosity model was validated in comparison with predictions using alternative approaches - in particular, of reverberation time ($T_{60}$) and steady-state sound-pressure level ($L_p$). The $T_{60}$ and $L_p$ values predicted by radiosity and diffuse-field theory, for a hypothetical room satisfying the requirements of a diffuse-sound field, were found to be equal within ±10% and ±0.2 dB, respectively. The $T_{60}$ values predicted by diffuse ray-tracing and radiosity models were also in good agreement, as expected. The $L_p$ values predicted by radiosity were slightly lower from those predicted by diffuse ray tracing. For more accurate $L_p$ results using radiosity, longer echogram lengths are required.

Echograms were also compared between the different prediction models. Echogram envelopes were similar for all prediction models, but the actual echogram representation varied significantly - in terms of the shape of the echogram and the number of zero-pressure amplitudes during the time investigated - between the different models. The echograms predicted by radiosity displayed a smooth, monotonic decay, unlike the echograms from other prediction models. The consequence of this smooth decay was a large energy loss at high frequencies when constructing the RIR. Possible explanations of this result are as follows:

1. The acoustical radiosity model intrinsically leads to inaccurate RIRs;
2. The method of converting echograms to a RIR is not valid for prediction models with smooth, monotonically-decaying echograms.
Chapter 9. Summary and Conclusions

The latter explanation suggests that fundamental digital-signal-processing requirements or basic acoustical theories are being violated. For a further understanding of the issues, future work could involve an evaluation of the existing methods used to convert echograms into RIRs and the possible development of a new method for this conversion. Comparing speech signals convolved with the predicted RIR, or even with the predicted echograms, with speech convolved with the measured RIR would also be useful.

The results above suggest that a new room-prediction model, which combines the method-of-images approach with radiosity, should predict valid $T_{60}$ and $L_p$ values, because these values are extracted from echograms, and not the RIR. Therefore, predicted results from the new room-prediction model were validated in comparison with measurements in the real rooms, by way of $T_{60s}$, $L_p$s, and echograms. Poor agreement between the measured and predicted $T_{60}$ and $L_p$ values were noted, with percentage deviation for $T_{60}$ and differences in $L_p$ of 45% and 10 dB, respectively. Discrepancies between the measured and predicted results were attributed to inaccurate values of physical parameters used in the predictions, as well as to the short echogram lengths.

To predict accurate $T_{60}$ and $L_p$ values with the new model, longer echogram lengths are required, with truncation times that are equivalent to or greater than the measured $T_{60}$ values. Furthermore, the transfer function of the loudspeaker and microphone should be completely compensated for in the model. Further work to improve the model related to the treatment of walls could involve the following:

- including the phase components of the wall's transfer function, which may have the secondary result of allowing accurate construction of the RIR;
- including the angular dependence of the incident wave on the absorption coefficient;
- increasing the frequency resolution of the wall's magnitude response, as opposed to using a constant value over a frequency range.

The radiosity component of the model could also be improved by implementing alternative, faster methods for form factor determination, such as the hemisphere or hemicube method.
A radiosity method that could include specular reflections while maintaining, or reducing, run times could also be developed. For a better representation of reality, the specular and diffuse models should be combined progressively as time advances, as opposed to the technique used in this thesis where the specular model predicts only the direct sound and first-order reflections, while the remainder is predicted using the radiosity model.

In summary, including the complex interactions between propagating waves and room surfaces make predicting the room sound field mathematically challenging. Various approaches have been used to predict room sound fields. These approaches are commonly validated physically. A more innovative approach is to validate the predictions subjectively using auralization. For real-time auralization processing, fast, efficient, and accurate models should be used to predict sound fields in rooms. With accurate sound-field prediction models, auralization becomes a powerful tool for hearing research and room design. In this thesis, a commercial auralization system was evaluated and a new sound-field prediction model was developed. With some refinements, the new sound-field prediction model, featuring radiosity, has the potential to become the foundation for future auralization applications.
- BIBLIOGRAPHY -


Bibliography


Bibliography


In this appendix, a script file containing the user input and program's responses, for an improved room sound-field prediction model, is shown. The program calculates the reverberation times, steady-state sound-pressure levels, and echograms at a receiver position. A description of most of the m-files are listed in Table 6.1. Those files that are not listed in Table 6.1 are documented in this appendix.

Script started on Mon Oct 6 12:39:14 1997
% matlab

MATLAB passcode expiration date of 01-dec-1997 is less than three months away.
UBC Electrical Engineering Matlab
Commands to get started: intro, demo, help help
Commands for more information: help, whatsnew, info, subscribe

Matlab 5.0 is also available; check out
http://www.ee.ubc.ca/local/software/matlab.html for details

>> element_size % user input: element_size.m asks the
% user for the dimensions of
% the room and creates
% vertices of the patches

Please give the dimensions of the room as [l w h] =
[3.9 5.3 2.7] % user input: length, width, and height
% of the room.

buffer =
Appendix A. Radiosity Room-Prediction Model

30
ans =
0.8333
buffer =
552
ans =
0.9583
buffer =
9120

>> radiosity % user input: calls ff_rev3.m and fullrad.m
Percent_done =
0
Percent_done =
1.0417
Percent_done =
2.0833
.
Percent_done =
96.8750
Percent_done =
97.9167
Percent_done =
98.9583

How many point sources are in the room? 1
Give ONE source location in metres as [length width height]: [2 4.86 1.29]
Give the power level for this source in Watts for each octave band:
[3.2e-4 6.33e-4 1.35e-3 2.25e-3 4.65e-3 5.71e-3]
percent_done = % This set of "percent_done" describes
1.0417 % the source energy propagating to
percent_done =
2.0833 % the different patches.
.
percent_done =
98.9583
percent_done = 100

percent_done = % this set of "percent_done" describes
0.0278 % the propagation of energy from one patch
percent_done = % to all other patches in an iterative
0.0556 % manner

percent_done = 99.9722
percent_done = 100

>> cd /tmp/wvaliani % changes the directory to /tmp/wvaliani
>> clear
>> load b_1000env % B-matrix for the 1000 Hz octave band

>> cd /a/frick/ext2a/year5/wvaliani/program
   % changes the directory to where the
   % variable.mat file is stored

>> load variable % variable.mat file
>> t_delay % user input: t_delay.m is responsible
   % for calculating the energy
   % at a receiver position

How many point receivers are in the room? 1

Give ONE receiver location in metres as [length width height]:
[1 2.36 1.29]

direct_wave = % direct wave energy in Watts
   0.0102
percent_done = 0.0262
percent_done = 0.0523
   .
   .

percent_done = 99.9738
percent_done = 100
SS_SPL = % the steady-state sound-pressure level
    85.4216 % for a given source and receiver position

RT_10 = 
    0.1521

RT_15 = 
    0.2168

RT_60 = % reverberation time as calculated from
    0.7770 % the T10 and T15 values.

>> exit

exit

script done on Mon Oct 6 15:47:49 1997

When the user inputs the command “radiosity”, as shown in the above script, the matlab
program calls a file, radiosity.m, and performs the necessary function of that file. The
contents of the radiosity.m file are shown below:

radiosity.m

path(path, '/a/frick/ext2a/year5/wvaliani/program')

load vertex.mat
    % contains the vertices for the room

area % determines the area of each patch and
    % the patch-patch distance

FF = ff_rev3(vertex, n_v, num_patch, PatchArea, mdpt,
    patch_dist);
    % determines the form factors for each
    % patch

save /a/frick/ext2a/year5/wvaliani/program/variable
    % saves the form factor matrix to a
    % *.mat file

B_save % an m-file that calls fullrad.m,
    % which calculates the radiant
Appendix A. Radiosity Room-Prediction Model

% intensity of each patch for all time

% this file also saves the B-matrix to the
% /tmp/wvaliani directory of the machine that
% matlab is currently running from

The file B_save.m, called within the file radiosity.m, calculates the radiant intensities for all patches for all time. The code for this program is shown below:

B_save.m

% calculates the impulse response of each patch in the room

load variable; % containing information about the
% patch areas, patch distances, and
% form factors

patch_abs_env; % an m-file containing information
% about the patch reflectivities

m_env; % an m-file listing the air absorption
% coefficients

source = input('How many point sources are in the room? '); % User inputs the number of sources in the room

for n = 1:source
    source_loc(n, 1:3) = input('Give ONE source location in
metres as [length width height]: '); % User inputs the number of sources in the room
    source_pow(n, 1:6) = input('Give the power level for this
source in Watts for each octave band: '); % User inputs the number of sources in the room
end

B = fullrad(p_125, FF, PatchArea, patch_dist, mdpt, m_125,
source, source_loc, source_pow(:, 1));
save /tmp/wvaliani/b_125env.mat B source_loc source source_pow

B = fullrad(p_250, FF, PatchArea, patch_dist, mdpt, m_250,
source, source_loc, source_pow(:, 2));
save /tmp/wvaliani/b_250env.mat B source_loc source source_pow

B = fullrad(p_500, FF, PatchArea, patch_dist, mdpt, m_500,
source, source_loc, source_pow(:, 3));
save /tmp/wvaliani/b_500env.mat B source_loc source source_pow
Appendix A. Radiosity Room-Prediction Model

B = fullrad(p_1000, FF, PatchArea, patch_dist, mdpt, m_1000, source, source_loc, source_pow(:, 4));
save /tmp/wvaliani/b_1000env.mat B source_loc source
source_pow

B = fullrad(p_2000, FF, PatchArea, patch_dist, mdpt, m_2000, source, source_loc, source_pow(:, 5));
save /tmp/wvaliani/b_2000env.mat B source_loc source
source_pow

B = fullrad(p_4000, FF, PatchArea, patch_dist, mdpt, m_4000, source, source_loc, source_pow(:, 6));
save /tmp/wvaliani/b_4000env.mat B source_loc source
source_pow

Samples of the Input Files patch_abs_env.m and m_env.m

patch_abs_env.m

% This file represents the reflection coefficients of the floor, wall and ceiling patches,
% respectively for a room containing 96 patches.

p_125 = [0.98 * ones(1, 16), 0.91 * ones(1, 64), 0.65 * ones(1, 16)];
p_250 = [0.98 * ones(1, 16), 0.94 * ones(1, 64), 0.78 * ones(1, 16)];
p_500 = [0.99 * ones(1, 16), 0.96 * ones(1, 64), 0.77 * ones(1, 16)];
p_1000 = [0.97 * ones(1, 16), 0.95 * ones(1, 64), 0.70 * ones(1, 16)];
p_2000 = [0.97 * ones(1, 16), 0.93 * ones(1, 64), 0.65 * ones(1, 16)];
p_4000 = [0.96 * ones(1, 16), 0.93 * ones(1, 64), 0.50 * ones(1, 16)];

m_env.m

% This file represents the octave-band air-absorption exponent for the room.

m_125 = [7.75*10^(-5)];
m_250 = [2.68*10^(-4)];
m_500 = [7.05*10^(-4)];
m_1000 = [1.32 * 10^(-3)];
m_2000 = [2.39 * 10^(-3)];
m_4000 = [5.94 * 10^(-3)];