Development of a Computationally Efficient Binaural Simulation for the Analysis of Structural Acoustic Data

Aimee L. Lalime

Thesis submitted to the Faculty of the Virginia Polytechnic Institute and State University In partial fulfillment of the requirements for the degree of

Master of Science In Mechanical Engineering

Dr. Marty E. Johnson, Chair Dr. Ricardo A. Burdisso Dr. Stephen A. Rizzi

August 12, 2002 Blacksburg, Virginia

Keywords: Virtual acoustics, HRTF, SVD, Equivalent source

Copyright 2002, Aimee L. Lalime
Binaural simulation is the recreation of a three-dimensional audio environment around a listener’s head. The binaural simulation of structural acoustic data would open new opportunities in virtual prototyping and simulation. By modeling the structure as an array of vibrating monopoles and applying Head Related Transfer Functions (HRTFs) to each of the sources, a binaural simulation of this type can be created. Unfortunately, this simulation method requires an extensive amount of computer power and speed for real-time simulation, more so than is available with current technology.

The objective of this research is to reduce the number of computations required in the binaural simulation of structural acoustic data. This thesis details two methods of reducing the number of real-time calculations required in this binaural analysis: singular value decomposition (SVD), and equivalent source reduction (ESR). The SVD method reduces the complexity of the HRTF computations by breaking the HRTFs into dominant singular values and vectors. The ESR method reduces the number of sources to be analyzed in real-time by replacing sources on the scale of a structural wavelength with sources on the scale of an acoustic wavelength. The ESR and SVD reduction methods can be combined to provide an estimated computation time reduction of 99.4%. In addition, preliminary tests show that there is a 97% correlation between the results of the combined reduction methods and the results found with current binaural simulation techniques.
Acknowledgements

I would like to thank my advisor, Marty, for teaching me acoustics, virtual acoustics, signal processing, and the value of checking my work. And, thanks for being understanding and laid-back, especially about my working from my laptop at home, Mill Mountain, or beside a lake 2000 miles away. I’m grateful to have worked with someone who is so easy to respect. Much the credit for this work is owed to you, because of your initiative and persistence.

A large portion of this work is also owed to others who supervised and assisted in this work: Steve Rizzi, Brenda Sullivan, Ferdinand Grosveld, and Bjoern Knoefel. In addition, thanks are due to Steve Rizzi and Ricardo Burdisso for being willing to serve on my committee.

Thanks, also, to all the folks at VAL, for your encouragement and friendship and for all the excellent work that you’ve done. Special thanks are also due to Michael Alley, for suggesting that I go to grad school, and to Lynda, Cathy, Ben Poe, and Jan Reiss for all of your help and encouragement.

I’d also like to thank my mother and father. You gave your children everything you ever had, the greatest of which are your love, encouragement, support, and the challenge to use our minds and gifts to pursue our dreams.

And this thesis is especially dedicated to my husband, Jay. It has been your character, devotion, encouragement, and inspiration that has enabled me to not only finish my Bachelor’s after getting married, but to also undertake a Master’s degree.

And finally, I want to thank God, for being my constant help.
Table of Contents

(Abstract) ................................................................................................................................. ii
Acknowledgements .................................................................................................................. iii
Table of Multimedia Objects ............................................................................................... vi
Nomenclature ......................................................................................................................... x
CHAPTER 1: INTRODUCTION ............................................................................................... 1
  1.1 Introduction to Binaural Acoustics ................................................................................ 1
  1.2 Motivation for Work ...................................................................................................... 4
  1.3 Review of Literature ...................................................................................................... 6
  1.4 Scope of Work ................................................................................................................ 10
CHAPTER 2: GENERAL THEORY OF BINAURAL ACOUSTICS ........................................... 12
  2.1 Plate Model .................................................................................................................... 12
      2.1.1 Theoretical development ..................................................................................... 12
      2.1.2 Plate response ..................................................................................................... 14
  2.2 Sound Radiation ............................................................................................................. 16
      2.2.1 Radiation from a monopole ................................................................................ 17
      2.2.2 Radiation from multiple sources ....................................................................... 18
      2.2.3 Radiation from a plate ........................................................................................ 19
  2.3 Binaural Calculations ................................................................................................... 21
  2.4 Exhaustive Method Schematic ...................................................................................... 24
  2.5 Real-time Binaural Simulation ....................................................................................... 25
  2.6 Conclusion ..................................................................................................................... 27
CHAPTER 3: REDUCTION METHODS .................................................................................. 29
  3.1 Singular Value Decomposition ..................................................................................... 29
      3.1.1 Description of SVD reduction method ................................................................. 29
      3.1.2 Schematic of SVD reduction method ................................................................ 32
      3.1.3 Number of important singular values ................................................................. 33
  3.2 Wavenumber Filtering .................................................................................................. 35
  3.3 Equivalent Source Reduction ......................................................................................... 37
      3.3.1 Description of Equivalent Source Reduction ..................................................... 38
Table of Multimedia Objects

Figure 1.1: Two sound rays approach the ears from a source ..........................................2
Figure 1.2: Knowles Electronic Manikin for Acoustic Research (KEMAR)^5 ....................3
Figure 1.3: The angles $\theta$ and $\phi$ as measured from the center of the head to the source.....3
Figure 1.4: Angles for which MIT’s HRTFs are defined and cataloged...........................4
Table 2.1: Parameters of plate model............................................................................14
Figure 2.1: Surface velocity of a plate excited by a point force ......................................15
Figure 2.2: Frequency response of the modeled plate.....................................................16
Figure 2.3: Pulsating sphere (monopole) radiating sound that is detected at point $R$ ......17
Figure 2.4: Three monopoles radiating sound to point $R$ .............................................19
Figure 2.5: A general planar structure radiating sound, which is detected at point $R$ ....20
Figure 2.6: The sound radiated to point $R$ by a vibrating plate.....................................21
Figure 2.7: (a) Pressure at the center of the head, and (b) pressure at the left ear ..........22
Figure 2.8: An array of monopoles (or vibrating plate) radiate sound to the left ear .....24
Figure 2.9: Schematic of the exhaustive method ...........................................................24
Figure 2.10: Schematic of a real-time binaural simulation.............................................27
Figure 3.1: (a) Contour plot of the HRIRs for $0^\circ$ - $360^\circ$ azimuth, and (b) same HRIRs for
only the $90^\circ$ and $-90^\circ$ azimuth cases........................................................................30
Figure 3.2: Matrix of HRIRs with ITD removed for each azimuth angle .......................31
Figure 3.3: SVD of the matrix of HRIRs of $0^\circ$ elevation.............................................31
Figure 3.4: Schematic of the SVD Method ......................................................................32
Figure 3.5: Singular value number verses magnitude for four frequency ranges ..........34
Figure 3.6: Error versus number of singular values for four frequency ranges ..........35
Figure 3.7: Wavenumber Decomposition of plate at 550 Hz ........................................36
Figure 3.8: Wavenumber filtering process ....................................................................37
Figure 3.9: Equivalent source reduction applied to a vibrating panel .........................40
Figure 3.10: Equivalent source reduction applied to a complex structure .................42
Figure 3.11: Equivalent source reduction applied to a vibrating enclosure ..............43
Figure 4.1: Schematic of the transmission loss facility^5 ...............................................46
Figure 4.2: Coordinate system used for data measurements^5 ....................................47
Figure 4.3: Mechanical excitation on the reverberant side of the panel

Figure 4.4: Three of the accelerometers attached to the aluminum panel

Table 4.1: Coordinates of measurement locations on aluminum test panel

Figure 4.5: Laser scanning vibrometer

Figure 4.6: Sum of the squares of the plate velocities

Table 4.2: Free field and reference microphone positions

Figure 4.7: (a) Transfer function between the input force and the microphone in the position shown in (b)

Figure 4.8: KEMAR mounted in support structure

Figure 4.9: (a) Transfer function from input force to left ear, configured as in (b)

Figure 4.10: (a) Consistency between two reference microphone measurements and (b) reference microphone orientation

Figure 4.11: Coherence of the transfer function from the input force to the reference microphone

Figure 4.12: Configuration of the aluminum test panel

Table 4.3: Test plate properties

Figure 4.13: Wavenumber decomposition of the plate accelerations at 550 Hz

Figure 4.14: Wavenumber decomposition at 2000 Hz (no aliasing)

Figure 4.15: Wavenumber decomposition at 4000 Hz (aliasing visible)

Figure 4.16: Frequency domain comparison of measured and calculated radiated pressures

Figure 4.17: Time domain comparison of the measured and calculated radiated pressures (extended view)

Figure 4.18: Time domain comparison of the measured and calculated radiated pressures (close-up view)

Figure 4.19: Head and plate configuration for HRTF application check

Figure 4.20: Frequency domain comparison of measured and calculated left binaural signal

Figure 4.21: Time domain comparison of the measured and calculated left binaural signal

Figure 4.22: Head and plate configuration for reduction method validations
Figure A2.4: Original image sources replaced with differently positioned image sources ................................................................. 101
Figure A4.1: Schematic of real-time binaural simulation using headphones ............... 105
Figure A5.1: HRIRs for 0° elevation and -180° to 180° azimuth .................................. 107
Figure A5.2: ITD found using the 5% rule (for 0° elevation HRIRs) ......................... 108
Figure A5.3: Polynomial fit of the ITD found using the 5% rule (for 0° elevation HRIRs)
......................................................................................................................................................................................... 109
Figure A6.1: Definition of wavelength 32 ................................................................ 111
Figure A6.2: Relationship between x, y and z components of the acoustic wavenumber
......................................................................................................................................................................................... 112
Figure A7.1: KEMAR mounted in the anechoic chamber with the baffled and reflective
panels .................................................................................................................................................................................. 115
Figure A7.2: KEMAR manikin and B-format microphone in front of the reflective and
baffled panels ................................................................................................................................................................... 116
Nomenclature
Abbreviations and symbols and equation variables

\( a \)  
Acceleration (m/s²)

\( a \)  
Complex acceleration vector

\( A_{mn} \)  
Complex resonance

B&K  
Bruel & Kjaer

C  
Conversion matrix

\( c \)  
Speed of sound in air (m/s)

\( C_E \)  
Computations required in the exhaustive or ESR method

\( c_{ph} \)  
Phase velocity of a wave (m/s)

\( C_S \)  
Computations required in the SVD method

D  
Bending stiffness (Nm)

\( E \)  
Modulus of Elasticity (N/m²)

ESR  
Equivalent Source Reduction

F  
Force amplitude (N)

H  
Matrix of HRIRs

\( h \)  
Plate thickness, in the \( z \) direction (m)

\( H_0 \)  
Matrix of speaker HRTFs

\( H_L \)  
HRIR for the left ear

\( H_R \)  
HRIR for the right ear

HRIR  
Head Related Impulse Response

HRTF  
Head Related Transfer Function

I  
Identity matrix

\( I \)  
Moment of Inertia (m³)

IID  
Interaural Intensity Difference

ITD  
Interaural Time Difference

\( k_a \)  
Acoustic wavenumber (rad/m)

\( k_{bxx} \)  
Bending wavenumber in the \( x \) direction (rad/m)

\( k_{bxy} \)  
Bending wavenumber in the \( y \) direction (rad/m)
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>KEMAR</td>
<td>Knowles Electronic Manikin for Electronic Research</td>
</tr>
<tr>
<td>KLE</td>
<td>Karhunen-Loève Expansion</td>
</tr>
<tr>
<td>$k_m$</td>
<td>Modal wavenumber in the $x$ direction (rad/m)</td>
</tr>
<tr>
<td>$k_n$</td>
<td>Modal wavenumber in the $y$ direction (rad/m)</td>
</tr>
<tr>
<td>$k_s$</td>
<td>Sample wavenumber</td>
</tr>
<tr>
<td>$L$</td>
<td>Number of equivalent sources</td>
</tr>
<tr>
<td>$L_x$</td>
<td>Plate dimension in the $x$ direction (m)</td>
</tr>
<tr>
<td>$L_y$</td>
<td>Plate dimension in the $y$ direction (m)</td>
</tr>
<tr>
<td>$M$</td>
<td>Number of singular values used to recreate the matrix of HRIRs</td>
</tr>
<tr>
<td>MIT</td>
<td>Massachusetts Institute of Technology</td>
</tr>
<tr>
<td>$N$</td>
<td>Number of sources</td>
</tr>
<tr>
<td>$N_E$</td>
<td>Number of equivalent sources</td>
</tr>
<tr>
<td>$p$</td>
<td>Sound pressure (Pa)</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal Components Analysis</td>
</tr>
<tr>
<td>$Q$</td>
<td>Volume velocity ($m^3$/s)</td>
</tr>
<tr>
<td>$R$</td>
<td>Observation point</td>
</tr>
<tr>
<td>$r$</td>
<td>Radial distance between the source and the measurement point (m)</td>
</tr>
<tr>
<td>$r_o$</td>
<td>Radius of the monopole (m)</td>
</tr>
<tr>
<td>$S$</td>
<td>Surface area ($m^2$)</td>
</tr>
<tr>
<td>SALT</td>
<td>Structural Acoustic Loads and Transmission</td>
</tr>
<tr>
<td>$S_E$</td>
<td>Evaluation surface</td>
</tr>
<tr>
<td>$S_O$</td>
<td>Original surface</td>
</tr>
<tr>
<td>$sr$</td>
<td>Sample rate (samples/sec)</td>
</tr>
<tr>
<td>SVD</td>
<td>Singular Value Decomposition</td>
</tr>
<tr>
<td>$T$</td>
<td>Matrix of complex transfer functions</td>
</tr>
<tr>
<td>$t_d$</td>
<td>Interaural time delay (samples)</td>
</tr>
<tr>
<td>TL</td>
<td>Transmission Loss</td>
</tr>
<tr>
<td>$u$</td>
<td>Complex velocity vector</td>
</tr>
<tr>
<td>$U$</td>
<td>Matrix of left singular vector</td>
</tr>
<tr>
<td>$u$</td>
<td>Velocity (m/s)</td>
</tr>
<tr>
<td>$U_i$</td>
<td>$i^{th}$ right singular vector</td>
</tr>
</tbody>
</table>
\( \mathbf{V} \) Matrix of right singular vectors
\( \mathbf{V}_i \) \( i^{th} \) left singular vector
\( \mathbf{W} \) Matrix of inverse transfer functions
\( W_{mn} \) Complex modal amplitude of the \( m^{th} \) and \( n^{th} \) mode (m/s)
\( x_f \) Position of point force in the \( x \) direction (m)
\( y_f \) Position of point force in the \( y \) direction (m)
\( \alpha \) Error coefficient
\( \beta \) Regularization parameter
\( \phi \) Elevation angle between the source and the center of the head
\( \varphi \) Angle of the sound radiated from the plate (rad)
\( \lambda \) Wavelength (m)
\( \theta \) Azimuth angle between the source and the center of the head
\( \rho_o \) Density of air (kg/m\(^3\))
\( \rho_s \) Density of the plate (kg/m\(^3\))
\( \Sigma \) Matrix of singular values
\( \sigma_i \) \( i^{th} \) singular value
\( \tau \) Temporal variable
\( \nu \) Poisson’s ratio of the plate
\( \omega \) Driving frequency of the force excitation (rad/s)
\( \omega_{mn} \) Natural frequency of the \( mn^{th} \) mode
\( \psi_{mn} \) Mode shape of the \( mn^{th} \) mode
\( \zeta_{mn} \) Damping ratio of the \( mn^{th} \) mode
CHAPTER 1: INTRODUCTION

The objective of this thesis is to communicate the author’s research in binaural acoustics, specifically in the creation of an efficient binaural simulation of structural acoustic data. In order to understand the motivation and scope of this research and the current literature on this topic, one must have an understanding of binaural acoustics, which will be described first.

1.1 Introduction to Binaural Acoustics

Binaural acoustics, three-dimensional sound, and virtual audio are three terms that are interchangeably used to describe the creation or recreation of a virtual sound source whose virtual location is different from that of the actual sound source.¹ ² For instance, sound played through headphones or stereo speakers can be digitally filtered such that the perceived location of that sound is not the actual headphones or speakers. Unfiltered sound from headphones or stereo speakers is termed monaural (not three-dimensional) because the perceived source location is the same as the actual source location. In addition, surround sound, while creating a fuller auditory experience, is also monaural sound because the perceived location of the sound source is the actual sound source location, even though the location and number of the speakers is different than in a typical stereo arrangement.

The virtual location of the sound source is achieved by applying signal processing to the monaural signal. First, the inverse of the acoustic transfer function from the speakers to the ears is applied to the monaural signal. Then, the measured acoustic transfer function from the new source position to the ears is applied to the signal. Through this process, the listener will sense that the sound is coming from the new source position rather than from the speakers.² For more information on this speaker arrangement, called the stereo dipole method, see Appendix 1.

The set of transfer functions that are used to create virtual 3-D sound are termed Head Related Transfer Functions (HRTFs). Contained within the HRTFs are two
primary localization cues and several lesser localization cues. The two most important cues are Interaural Intensity Difference (IID) and Interaural Time Difference (ITD)\textsuperscript{1,3,4}, IID is the difference in sound intensity that is detected by each ear and ITD is the difference in the amount of time it takes for the sound to travel from the source to each ear. Figure 1.1 shows two rays of sound emitted from the same source and traveling to the right and left ears. In this example, the sound at the left ear will be heard slightly sooner and will be louder than the sound heard at the right ear.

![Figure 1.1: Two sound rays approach the ears from a source](image)

HRTFs also contain localization cues that describe how sound reflects off the human torso, wraps around the head, and resonates in the inner and outer ear. Since every human torso, head, and ears are different, each person has their own individual HRTFs. A generalized set of HRTFs can be obtained using the Knowles Electronic Manikin for Acoustic Research (KEMAR). The KEMAR manikin (figure 1.2) has a head, torso, and ears that are specifically shaped in order to represent the average human form. The set of generalized HRTFs are obtained by taking the transfer function between the source acceleration and the pressure detected at two microphones inside each of the KEMAR’s ears.
Technically, HRTFs are functions of frequency (or time), angle, and distance. In the time domain, an HRTF is referred to as a Head Related Impulse Response (HRIR), which represents the impulse response between a source of sound and a human’s ears. The angles corresponding to each HRTF are designated $\theta$ and $\phi$ and represent the azimuth and elevation angles, respectively, of the source (figure 1.3). The angles are measured between the center of the head and the source, with $0^\circ$ elevation being defined as the horizontal plane that intersects the two ears and $0^\circ$ azimuth being defined as the vertical plane that intersects the center of the head and the nose.

Because the researchers use HRTFs that were found using several sound sources placed at a uniform distance from the head, these measured HRTFs are not actually a
function of distance. The sound attenuation and delay due to distance are included in binaural simulation by implementing Rayleigh’s equation. That is, the radiation of sound from the source to the head is calculated or measured prior to the addition of the HRTFs.

In this analysis the authors used a set of generalized HRTFs that were recorded at the ears of a KEMAR dummy head (Pinnae type DB-065) by researchers at MIT. The HRTFs in the MIT set are defined from -40° to 90° with an elevation step of 10° and have a sample rate of 44.1 kHz. Since the HRTFs are defined spherically, the degree step in the circumferential direction is different for each elevation. Figure 1.4 shows the different circumferential angle steps for each elevation. Note that the circumferential angle steps are in purple and the elevations are in red. Further note that the MIT HRTFs are not a minimum phase representation, that is, the delay between the right and left ears are contained in the HRTF itself, rather than in a delay look-up table.

![Figure 1.4: Angles for which MIT's HRTFs are defined and cataloged](image)

1.2 Motivation for Work

The binaural technology described in the previous section is currently used to create virtual simulations for a small number of point sources. The expansion of this
technology to represent multiple sources or a distributed source would open up new possibilities in virtual simulation and virtual prototyping. For example, a virtual reality simulation of the inside of the space station could be greatly enhanced by incorporating the three-dimensional sound produced by the vibrating walls. Fully spatialized audio would allow a more immersive three-dimensional simulation, allowing astronauts to experience and prepare for life in the space station in a much less expensive, error-friendly environment.

Another application for a binaural simulation of a vibrating structure is in the area of virtual prototyping. Virtual prototyping uses virtual reality technology to view and analyze prototypes of automobiles, airplanes, buildings, etc. This concept of virtual prototyping can be extended to analyze the acoustical properties of a prototype. For example, a three-dimensional audio-visual representation of a concert hall would allow architects and analysts to experientially determine the acoustic characteristics of the room. Incorporating the visual aspect with the acoustic allows the user to sift rapidly through noise data to determine what parts of a vibrating structure are radiating sound to what areas in an acoustic enclosure. This acoustical virtual prototyping would also be useful in the design of speakers and musical instruments.

Unfortunately, the real-time binaural simulation of structural data is not currently feasible because it requires more computer speed and power than is available with modern technology. The intuitive method for creating a binaural simulation of this type is to model the structure as an array of vibrating monopoles, apply the HRTFs to each of these sources, and sum the resulting signals. However, thousands of monopoles are required to accurately represent the characteristics of a vibrating wall, meaning that thousands of convolutions must be carried out per iteration. For a sample rate of 44.1 kHz, each iteration must be carried out in less than 23 µs.

Since it is currently impossible to perform thousands of convolutions in 23 µs, the objective of the author’s research has been to reduce the number of calculations required in order to determine the binaural signals of a vibrating structure. Two main reduction methods were investigated: (1) Equivalent Source Reduction (ESR), a pre-processing method which reduces the number of monopoles that are needed in order to represent the
vibrating source, and (2) Singular Value Decomposition (SVD), a real-time method which replaces the convolution with faster operations.

1.3 Review of Literature

The research outlined in this thesis combines ESR and SVD and applies them to the binaural simulation of a vibrating structure. While the binaural simulation of a vibrating structure has not previously been researched, the modeling of HRTFs has been researched extensively. Since the SVD method models the HRTFs as a weighted sum of invariant singular vectors, similar research in modeling HRTFs will be discussed here. Current research in equivalent source reduction will also be discussed.

Researchers have attempted to model HRTFs, or create a functional representation of the HRTFs, for several reasons. According to Cheng and Wakefield’s invited tutorial on HRTFs³, one motivation for modeling HRTFs has been to reduce errors in virtual acoustic displays. Because of the noise inherent in digital signal processing, sounds sometimes seem to come from inside the head rather than from outside. Another error is front-back confusion, in which virtual sources in front of the listener are confused with virtual sources behind the listener, and vice versa. In addition, sounds from elevations other than zero are difficult to simulate effectively. Modeling the HRTFs in terms of acoustically relevant components allows the researcher to understand the different spectral cues and over-emphasize them in the HRTF (above the noise) in order to minimize front-back confusion, etc.

Another reason to model HRTFs is to reduce the number of HRTFs that must be measured for a particular individual. If the HRTFs could be defined as a mathematical function of source position, one would only need to measure one HRTF per individual at any arbitrary angular position. This single measured HRTF could be extrapolated (using the mathematical function for angular dependence of HRTFs) in order to functionally represent a set of HRTFs describing the infinite number of angular positions surrounding the individual’s head. Such a model of the HRTFs would also remove the need for interpolation between listener and source positions in a virtual acoustic display. (The current need for interpolation is due to the fact that each set of HRTFs is only defined for
a finite number of source positions.) This functional representation would also occupy less storage space than a full set of HRTFs.

Providing a functional representation of HRTFs can also achieve a reduction in the number of computations required in binaural simulation. As mentioned previously, the process of convolving the monaural signal with its appropriate HRTF is the most time consuming computational process in binaural simulation. If a low-order approximation of the HRTFs can be made, replacing this HRTF model into the binaural simulation may yield great computational reduction.

The first research into modeling HRTFs was performed by Martens. Martens used a principal components analysis (PCA) to determine the directional cues found in HRTFs. PCA is an orthogonal matrix decomposition that is performed on the log of the magnitude of the HRTFs. The HRTFs can then be represented as a weighted sum of basis functions, each of which are assumed to account for different directional cues. HRTFs for 36 different directions in the horizontal plane were considered. While Martens’ work focused only on azimuth cues, Kistler and Wightman confirmed Martens’ results and performed PCA on HRTFs of different elevations as well. This work performed PCA on a group of 5300 HRTFs that were gleaned from 256 different directions and from 10 different subjects. Their work was also perceptually validated using human subjects. Since the principal components found with PCA can vary across laboratories, Middlebrooks and Green performed an independent study performing PCA on HRTFs measured using slightly different methods and were able to verify the principal components found by Kistler and Wightman.

Another term for PCA is Karhunen-Loève expansion (KLE). Although the mathematics behind PCA and KLE are exactly the same, different results are provided by Chen because he performed KLE on the complex-valued HRTFs rather than on the log of the magnitude of the HRTFs. In addition, regularization theory was used to create a continuous functional representation of HRTFs in terms of weighted eigenfunctions (basis functions). Although computational reduction is mentioned as a potential benefit of this spatial feature extraction and regularization, the driving philosophy is functional approximation. This functional approximation seeks to solve three problems of measured HRTFs: (1) measuring a set of HRTFs on a sufficiently fine grid requires long data
collection times, (2) it is difficult to visualize the characteristics of such a large data set, (3) the measured HRTFs represent only discrete samples of a continuous auditory space. The results of this functional approximation were validated by comparing approximate and measured HRTFs for a human and a live anesthetized cat. This KLE research has also been performed in the time domain and validated using cat HRTFs.11

Singular value decomposition (SVD) can also be used to model HRTFs, a methodology which was first defined by Abel and Foster.12, 13 The HRTFs are broken into a sum of weighted filters, just as in the PCA and KLE cases. In addition, the weighting coefficients (singular values) are the square root of the eigenvalues of Karhunen-Loève expansion. However, SVD is a more streamlined mathematical operation because it does not require the computation of a covariant matrix. Abel and Foster also defined a method by which multiple inputs and multiple outputs can be more efficiently processed in binaural simulation. They estimated that twelve to sixteen filters (or singular values) are required in order to accurately reproduce the HRTFs. However, their research did not provide analytical results to prove the accuracy or computational effectiveness of the SVD method. In contrast, the author’s research uses measured data to prove that the SVD method provides a significant computational reduction while maintaining the accuracy of current methods involving measured HRTFs.

HRTFs can also be represented as a weighted sum of surface spherical harmonics.14 Although modeling HRTFs in terms of spherical harmonics offers no computational advantages, one can gain insight into the directional cues embedded within the structure of the HRTFs. Because spherical harmonics are a well-known and hierarchical set of orthogonal basis functions, the directional cues are easier to extract from the HRTFs.

In addition, Nelson and Kahana have been able to draw connections between spherical harmonics and the SVD of a set of Green’s functions between a radiating sphere and the pressure at a set of free-field points of uniform radius from the sphere.15 In particular, they have found that the resulting left and right singular vectors are related to the spherical harmonics by a unitary transformation. The SVD of the Green’s functions of a radiating ellipsoid and pinnae were also studied, assumedly with the goal of eventually relating the SVD of HRTFs to classical spherical harmonics. This would
imply that the left and right singular vectors found by performing the SVD of HRTFs provide spatial information as well as mathematical efficiency. Unfortunately, the source and field must be sampled on a fine mesh in order to relate the singular vectors to spherical harmonics. Because of this limitation, the results from the author’s current research does not provide any significant information concerning the directional cues that comprise HRTFs. However, future work in this area could yield promising conclusions regarding the structure of HRTFs.

Other research into modeling HRTFs has included using pole-zero approximations 16-18, state space models19, 20, balanced model truncation21, and structural modeling22, 23. These methods are each different from the authors’ in that they do not provide a mathematical model of the HRTFs. Each method has its own advantages, though. The pole-zero method reduces the dimensionality of the measured set of HRTFs (thereby reducing the computational load) and creates filters that do not need to be updated based on listener position. Although the use of a state-space model does not provide a computational reduction over current simulation methods, it is advantageous because it eliminates the need for HRTF interpolation (which introduces more error into the virtual acoustic display) and switching of HRTFs as the listener’s position changes. Switching HRTFs can lead to the “clicking sound” that is sometimes found in binaural simulation. Balanced model truncation replaces the FIR filters with IIR filters, producing a computational reduction. In addition, a structural model developed by Brown and Duda models the HRTFs in terms of four mathematical functions, which are head shadow and delay, shoulder echo, room echo, and pinnae echoes. This simple model offers computational reduction, but is not fully developed and therefore is not yet an accurate reproduction of the HRTFs. It should be noted that of the above methods that offer a computational reduction, only the structural model offers a quantitative measure of that reduction. The structural model reduces the number of multiplications per HRIR from 128 to 35. Duda offers a good summary of the different HRTF modeling methods.24

In addition to modeling the HRTFs using singular value decomposition, equivalent source reduction (ESR) has also been investigated as a reduction method. This method was first developed by Koopman et. al. and allows the radiation from a source in a free field to be efficiently calculated.25 Other research in this area has
investigated the robustness of the ESR method \cite{26-29}, developed a time domain formulation of this method\cite{30}, and developed the ESR technique for calculating the sound field inside an enclosure\cite{31}.

While much research has already been performed on the modeling of HRTFs and on ESR, the current research provides several contributions to the current technology of binaural simulation and analysis of structural acoustic data. The SVD method is applied to a vibro-acoustic source, which has not yet been discussed in current literature. Similarly, the ESR method developed by Koopman and others is applied to a binaural simulation. After using measured structural acoustic data to implement the ESR and SVD methods in binaural simulation, the results are quantitatively compared with the exhaustive method of binaural simulation. Finally, the number of computations involved in each method are estimated and compared.

1.4 Scope of Work

As mentioned previously, the goal of the current research has been to use ESR and SVD to reduce the number of computations required in the real-time binaural simulation of a vibro-acoustic source. Since the principal restriction in creating this real-time simulation is the extensive number of computations that it requires, the goal of this research is primarily to reduce the computational load, rather than to better understand the HRTFs, to eliminate the need for interpolation between listener positions, to reduce errors in binaural simulation due to measured HRTFs, or to reduce the number of measurements required in order to define a set of individualized HRTFs.

In addition, the current research deals only with a measured set of generalized HRTFs, rather than with individualized ones (although all the research performed here can easily be applied to any individualized set). Similarly, the test case used to verify the reduction methods is that of a vibrating panel radiating sound into an anechoic chamber. Because of this arrangement, reflection and room acoustics are not thoroughly addressed. In theory, the reflections inside the room can be modeled in terms of image sources (see Appendix 2) and equivalent sources can replace both direct and image sources, but validating this theory with results is a matter of future work.
The second chapter of this thesis will explain the theory behind binaural acoustics and vibrating structures and develop the exhaustive method of binaural simulation. This discussion of theory will be concluded in the third chapter with a description of the theory behind the reduction methods that were investigated, most notably SVD and ESR. The fourth chapter will depict the implementation of the simulation methods using measured structural acoustic data. Finally, conclusions will be drawn regarding the effectiveness of the reduction methods and future work in this research area will be discussed.
CHAPTER 2: GENERAL THEORY OF BINAURAL ACOUSTICS

In order to discuss the reduced methods of simulating the binaural signals associated with structural acoustics, there are several topics that must first be understood. The first step in the author’s research was to create a modal model of a vibrating plate in order to begin the binaural simulation process, which is illustrated in section 2.1. Later in the research, the results from this model were replaced with measurements taken by Grosveld. Following an account of this plate model will be the theory behind radiation of sound from a plate into the free field. Since the plate is modeled as a group of vibrating monopoles, the theory behind radiation from a monopole is also introduced in section 2.2. Section 2.3 will describe the binaural calculations involved in the “exhaustive” method of binaural simulation and section 2.4 will introduce the exhaustive method schematic. This method is termed the exhaustive method because it is computationally expensive compared with the reduced methods, which will be depicted in chapter 3. Finally, section 2.5 will discuss other elements of real-time binaural simulation, including head tracking and the stereo-dipole method.

2.1 Plate Model

In order to develop and test a binaural simulation of structural acoustic data, one must begin with measured or calculated plate accelerations. In this research, measured plate data was ultimately used, but prior to the collection of those measurements, a model of a vibrating plate was computed in Matlab using standard plate theory. The modeled plate was a simply supported plate in an infinite baffle, which was excited by a point force and radiated sound into the free field. The in-vacuo response of the plate was found and used to develop the exhaustive method of binaural simulation.

2.1.1 Theoretical development

Modal analysis was used to describe the plate vibration as a finite sum of structural modes. Modes can be considered to be standing waves that are formed by the
interaction of several reflected traveling waves. The velocity, \( u \), normal to the plane of the plate is defined according to equation 2.1:

\[
 u(x, y) = \sum_{m=1}^{M} \sum_{n=1}^{N} W_{mn} \psi_{mn}(x, y)
\]

where \( W_{mn} \) is the complex modal amplitude of the \( m^{th} \) and \( n^{th} \) mode, and \( \psi_{mn} \) is the mode shape of the \( mn^{th} \) mode, defined by:

\[
 \psi_{mn}(x, y) = \sin(k_m x)\sin(k_n y)
\]

where \( k_m = m\pi/L_x \) and \( k_n = n\pi/L_y \) are the modal wavenumbers in the \( x \) and \( y \) directions, respectively. \( (L_x \text{ and } L_y \text{ are the dimensions of the plate in the } x \text{ and } y \text{ directions.}) \) Modes up to \( m = M \) and \( n = N \) are considered in the finite modal summation.

The modal amplitude \( W_{mn} \) is dependent upon the amplitude, \( F \), and location, \( x = x_f \) and \( y = y_f \), of the point force, and is defined such that:

\[
 W_{mn}(x, y) = \frac{4F \psi_{mn}(x_f, y_f)}{\rho_s h L_x L_y} A_{mn}(\omega)
\]

where \( \rho_s \) is the density of air, \( h \) is the plate thickness, and \( A_{mn}(\omega) \) is the complex resonance term, which is defined by:

\[
 A_{mn}(\omega) = \frac{\omega}{2\zeta_{mn}\omega_{mn}} \left( \frac{\omega_{mn}^2 - \omega^2}{\omega_{mn}^2 - \omega^2} \right)
\]

where \( \omega \) is the driving frequency of the force excitation, \( \zeta_{mn} \) is the damping ratio of the \( mn^{th} \) mode, and \( \omega_{mn} \) is the natural frequency of the \( mn^{th} \) mode, and is given by:

\[
 \omega_{mn} = \left( \frac{EI}{\rho_s h} \right)^{1/2} \left( k_m^2 + k_n^2 \right)
\]

where \( E \) is the modulus of elasticity and \( I \) is the moment of inertia of the plate, defined as:

\[
 I = \frac{h^3}{12(1 - \nu^2)}
\]

where \( \nu \) is the Poisson’s ratio of the plate. The velocity of the plate, then, as a function of position \( x, y \), can be summarized to be:

\[
 u(x, y, \omega) = \frac{4F}{\rho_s h L_x L_y} \sum_{m=1}^{M} \sum_{n=1}^{N} A_{mn}(\omega) \psi_{mn}(x, y) \psi_{mn}(x_f, y_f)
\]
This continuous function can be used to create a finite set of discrete plate velocities, which can represent the characteristics of the continuous plate.

2.1.2 Plate response

When the modal summation is used to model an aluminum plate (with the plate parameters shown in table 2.1), the modal response corresponding to a driving frequency of 100 Hz is found according to figure 2.1.

<table>
<thead>
<tr>
<th>Plate Parameter</th>
<th>Numerical Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Damping Ratio, $\zeta$</td>
<td>0.01</td>
</tr>
<tr>
<td>Force position in x direction, $x_f$</td>
<td>0.2 m</td>
</tr>
<tr>
<td>Force position in y direction, $y_f$</td>
<td>0.35 m</td>
</tr>
<tr>
<td>Modulus of elasticity, $E$</td>
<td>$7.1 \times 10^{10}$ N/m$^2$</td>
</tr>
<tr>
<td>Plate density, $\rho_s$</td>
<td>2700 kg/m$^3$</td>
</tr>
<tr>
<td>Plate dimension in x direction, $L_x$</td>
<td>0.50 m</td>
</tr>
<tr>
<td>Plate dimension in y direction, $L_y$</td>
<td>0.75 m</td>
</tr>
<tr>
<td>Plate dimension in z direction, $h$</td>
<td>0.005 m</td>
</tr>
<tr>
<td>Poisson’s ratio, $\nu$</td>
<td>0.33</td>
</tr>
</tbody>
</table>
The frequency response of the same plate (parameters of table 2.1) is shown in figure 2.2. In this case, the surface velocity is calculated at $x = 0.3$ and $y = 0.7$ for driving frequencies ranging from 0 to 4000 Hz. In this modal summation, 20 modes were considered in each of the $x$ and $y$ directions (that is, $M$ and $N$ of equation 2.7 were both set to 20). Notice that up to about 1000 Hz, very sharp resonance peaks are visible in the modeled plate response, which is normal plate behavior.
2.2 Sound Radiation

In order to determine the binaural signals that will represent the plate (or any other planar structure) one must determine the sound at the center of the listener’s head (if the head were not present) and then apply the appropriate HRTFs. This is because the MIT HRTFs do not provide distance cues (as mentioned in section 1.1). The sound at the center of the listener’s head can be found by modeling the plate as an array of monopoles (each having uniform directivity). This is advantageous because the discrete set of calculated (or measured) velocities on the surface of the plate can be assumed to be the velocities of replacement monopoles in the same positions. Using Rayleigh’s equation, the radiated sound pressure due to that monopole can then be found. For multiple monopoles, the sound signals due to each monopole can be summed to result in the overall pressure. Similarly, a plate or general planar structure can be represented by summing the radiated sound pressure due to the contribution of a planar array of
monopoles. These types of sound radiation will be discussed in more detail in the following sections. This development will be presented in the time domain and in the frequency domain because the calculation of binaural signals can be performed in either domain. However, real-time implementation requires that the process be performed in the time domain, so the actual implementation was in the time domain.

2.2.1 Radiation from a monopole

The simplest acoustic source is that of a pulsating sphere, also known as a monopole (figure 2.3). The radius of the monopole, \( r_o \), is exaggerated in figure 2.3 for clarity, but is assumed to be very small compared with the acoustic wavelength. According to Fahy\(^{32,33}\), the sound pressure, \( p \), due to a monopole is defined in the time domain according to equation 2.8:

\[
p(r, t) = \left( \frac{\rho_o}{4\pi r} \right) \frac{\partial}{\partial t} \left[ Q \left( t - \frac{r}{c} \right) \right]
\]

and in the frequency domain\(^{32}\) as:

\[
p(r, \omega) = j\omega \left( \frac{\rho_o}{4\pi r} \right) Q e^{j(\omega t - k_ar)}
\]

where \( r \) is the radial distance from the monopole to the observation point (termed point \( R \)), \( \rho_o \) is the density of air, \( Q \) is the volume velocity, \( c \) is the speed of sound in air, and \( k_a \) is the acoustic wavenumber, \( \omega c.\) \(^{33}\) \( Q \) is defined according to equation 2.10:

\[
Q = Su
\]

where \( S \) is the surface area of the monopole \((4\pi r_o^2)\) and \( u \) is the velocity of the surface of the monopole.

![Figure 2.3: Pulsating sphere (monopole) radiating sound that is detected at point R](image)
In the case of a baffled vibrating hemisphere rather than a vibrating sphere, the pressure at the radial distance \( r \) will be:

\[
p(r, t) = \left( \frac{\rho_0}{2\pi r} \right) \frac{\partial}{\partial t} \left[ Q \left( t - \frac{r}{c} \right) \right] = \left( \frac{\rho_0}{2\pi} \right) B \left( t - \frac{r}{c} \right)
\]  

where \( B \) is the volume acceleration (\( Sa \)). In the frequency domain, equation 2.11 becomes:

\[
p(r, \omega) = \left( \frac{\rho_0}{2\pi} \right) B e^{i(\omega - kr)}
\]  

where \( a \) is the acceleration of the surface of the pulsating sphere, \( \frac{du}{dt} \) (figure 2.3). The relationship of equations 2.11 and 2.12 signifies that the sound at point \( R \) will have a magnitude that varies inversely with the distance \( r \) and a time delay equal to the time it takes for the sound to reach point \( R \), which is \( r/c \).

### 2.2.2 Radiation from multiple sources

Having determined the pressure at point \( R \) due to a single monopole source, the pressure resulting from a group of monopoles can be calculated as a linear superposition of the pressure due to each source. Equations 2.13 and 2.14 show the total pressure, \( p(t) \) or \( p(\omega) \), at point \( R \) resulting from \( N \) monopole sources, each of which has a volume velocity \( B_n \) which is dependent upon the surface area and acceleration of the \( n^{th} \) monopole:

\[
p(t) = \sum_{n=1}^{N} p_n(r_n, t) = \sum_{n=1}^{N} \left( \frac{\rho_0}{2\pi r_n} \right) B_n \left( t - \frac{r_n}{c} \right)
\]  

\[
p(\omega) = \sum_{n=1}^{N} p_n(r_n, \omega) = \sum_{n=1}^{N} \left( \frac{\rho_0}{2\pi r_n} \right) B_n e^{i(\omega - kr_n)}
\]  

where \( r_n \) represents the distance between the \( n^{th} \) source and the point \( R \).
An example of this is shown below in figure 2.4. The pressure at the microphone due to each of three monopole sources is summed to equal $p$. This linear superposition process can be applied to any number of sources.

![Figure 2.4: Three monopoles radiating sound to point R](image)

2.2.3 Radiation from a plate

It is a common practice among acousticians to replace a vibrating planar structure with an array of baffled vibrating hemispheres.\textsuperscript{32,33} Although this is an approximation, it is very accurate if the spacing of the sources is less than half the structural and acoustic wavelengths. Figure 2.5 shows the sound radiated from a planar structure (or array of monopoles) into the free field and detected by a microphone at point $R$, where $\theta_n$ and $\phi_n$ represent the azimuth and elevation angles between the $n^{th}$ monopole and point $R$. 
The goal of this research is to find efficient ways to analyze structural acoustic data through binaural simulation. To develop these efficient methods, the authors have chosen to focus on the analysis of a vibrating plate. This plate is set in an infinite baffle and radiates sound into the free field. This structure and configuration was chosen for two reasons: 1) the plate is a relatively simple structure to analyze, allowing the research to focus on the binaural processes rather than structural effects, and 2) a plate in this configuration is essentially the same as a plate secured in a baffle, radiating sound into an anechoic chamber, which can be arranged in a typical acoustic laboratory.
2.3 Binaural Calculations

Having determined the monaural signals due to a vibrating structure, it remains to apply HRTFs to determine the binaural signals. Similar to the sound radiation development, the application of HRTFs to monaural signals can be performed in either the time or frequency domain. The equations to follow will be presented in both domains.

As mentioned in section 1.1, HRTFs are a set of acoustic transfer functions that describe how sound is reflected off the human torso, wraps around the head, and resonates in the inner and outer ear. Because the human body is not axisymmetric, the HRTFs vary depending upon the elevation angle, $\phi$, and the circumferential angle, $\theta$, between the sound source and the center of the human head (see figures 1.3 and 1.4). Since every human torso, head, and ears are different, each person has their own individual HRTFs. Generalized HRTFs, while less convincing than individual ones, can also be effectively used in binaural simulation.\textsuperscript{34,35} This research uses generalized...
HRTFs that were measured by MIT\textsuperscript{6} using a Knowles Electronic Manikin for Acoustic Research (KEMAR). However, all of the work presented here can easily be applied to any set of individualized HRTFs.

The binaural signals (for the left and right ears) can be found by convolving the monaural signal with the appropriate Head Related Impulse Response (HRIR), which is the time domain version of the HRTF. However, since the HRIRs depend on direction, a binaural signal that accurately represents the vibrating structure cannot be found by convolving the monaural signal ($p(t)$ from equation 2.13 or $p(\omega)$ from 2.14) with a single HRIR. Instead, the signal from each monopole must be convolved with its corresponding HRIR to find its contribution to the sound pressure at the left ear (figure 2.7b). For example, $p_{Ln}$ is the sound pressure at the left ear due to the $n^{th}$ monopole.

\begin{equation}
    p_{Ln}(t) = \int_{0}^{\infty} H_L(\theta_n, \phi_n, \tau) p_n(t-\tau) d\tau
\end{equation}

and in the frequency domain, convolution becomes multiplication:

\begin{equation}
    p_{Ln}(\omega) = H_L(\theta_n, \phi_n, \omega) p_n(\omega)
\end{equation}

where $H_L$ is the HRIR for the left ear, $\tau$ is a temporal variable, and $p_{Ln}(\omega)$ and $p_n(\omega)$ are complex values. The pressure, $p_n$, resulting from the $n^{th}$ monopole at angles $\theta, \phi$ with respect to the head, represents the pressure at the center of the head if the head were not present (figure 2.7a). A similar equation can be written to describe the response at the right ear.

Figure 2.7: (a) Pressure at the center of the head, and (b) pressure at the left ear
Since the binaural signals are calculated using a computer, the convolution of equation 2.15 is performed on discrete sampled signals rather than continuous ones.

\[ p_{Ln}[i] = \sum_{j=0}^{M-1} H_L[\theta_n, \phi_n, j] p_n[i-j] \]  

(2.17)

The variable \( M \) is the number of samples used in the HRIR FIR filter and \( i \) is the \( i^{th} \) discrete time step. Similarly, the frequency domain equation 2.16 becomes:

\[ p_{Ln}[\omega] = H_L[\theta_n, \phi_n, \omega] p_n[\omega] \]  

(2.18)

The resulting binaural signal, \( p_L \), from the vibrating structure is found by summing the pressure contributions from each monopole:

\[ p_L[i] = \sum_{n=1}^{N} p_{Ln}[i] = \sum_{n=1}^{N} \sum_{j=0}^{M-1} H_L[\theta_n, \phi_n, j] p_n[i-j] \]  

(2.19)

or in the frequency domain:

\[ p_L[\omega] = \sum_{n=1}^{N} p_{Ln}[\omega] = \sum_{n=1}^{N} \sum_{j=0}^{M-1} H_L[\theta_n, \phi_n, \omega] p_n[\omega] \]  

(2.20)

Equations 2.19 and 2.20 will calculate the sound in the left ear due to a vibrating plate, a configuration that is shown in figure 2.8.
2.4 Exhaustive Method Schematic

Figure 2.9 is a basic schematic of the exhaustive method of calculating the binaural signals due to a vibrating structure for a static listener (i.e., the listener does not move in real-time with respect to the vibrating structure). This process can be performed in both the time and frequency domains, but the real-time simulation (which will be discussed in section 2.4) requires that calculations be performed in the time domain. Because the purpose of this calculation is for inclusion in a real-time binaural simulation, this process will be described in the time domain.

The inputs into this mathematical process are the accelerations of each of the $N$ monopole sources. These accelerations are calculated or measured a priori. The radiation model (equation 2.11) applies a time delay and gain to each monopole’s acceleration in order to define the pressure at the center of the head due to each of the $N$ monopoles. Since equation 2.11 depends upon the radial distance, $r_n$, between the center of the head and the source, the radiation model applies a different time delay and gain to the accelerations from each of the $N$ monopoles (denoted by the arrow through the radiation model square). This radial distance, $r_n$, in spherical coordinates, is found according to equation 2.21.

$$r_n = \sqrt{(x_n - x_h)^2 + (y_n - y_h)^2 + (z_n - z_h)^2}$$  \hspace{1cm} (2.21)

where $x_n$, $y_n$, and $z_n$ are the Cartesian coordinates of the $n^{th}$ monopole and $x_h$, $y_h$, and $z_h$ are the Cartesian coordinates of the center of the listener’s head. Unless a head-tracking unit is included in the binaural simulation (see section 2.5), the variables $x_h$, $y_h$, and $z_h$ will be constant valued.
As in equation 2.17, the 128 coefficient HRIRs for the left ear are convolved with the \( N \) pressures at the head center to get the pressure at the left ear due to each of the \( N \) monopoles. The HRIR filters (cataloged for various discrete azimuth and elevation angles) are stored in a look-up table and the appropriate HRIR is chosen based upon the azimuth and elevation angles, \( \theta_n, \phi_n \), between the listener and the \( n^{th} \) monopole.

Note that the head has six degrees of freedom. That is, its orientation is designated by its \( x_h, y_h, \) and \( z_h \) position and its \( \theta_h, \phi_h, \) and \( \xi_h \) angular rotation. As before, \( \theta_h \) and \( \phi_h \) designate the azimuth and elevation angular rotation of the head (that is, rotation about the \( y \) and \( x \) axes, respectively), and \( \xi_h \) refers to the head rotation about the \( z \) axis. However, the position of each monopole source relative to the head can be denoted by only three spherical coordinates, \( r_n, \theta_n, \) and \( \phi_n \), where \( \theta_n \) and \( \phi_n \) are functions of the position and angular rotation of the head and the position of the monopole source.

The pressures due to each monopole source are then summed to obtain the binaural signal for the left ear (the parallel process must also be performed for the right ear). If there were no error in the measurement or computer processes, the resulting binaural signals would be the same as those measured inside the ears of someone listening to a structure radiating sound into the free field. It should be noted that this method assumes that the listener does not affect the radiation from the sources and is essentially acoustically invisible.

The most computationally expensive calculations in the binaural simulation are the binaural calculations discussed in this section. Specifically, when \( N \) is large, the convolution of the \( N \) sources with the appropriate HRTF filters requires an excessive amount of computation time. The reduction methods discussed in chapter 3 will focus on reducing the computations involved in this part of the binaural simulation by reducing either the number of input sources or replacing the convolution with faster operations.

### 2.5 Real-time Binaural Simulation

The binaural calculations presented in the previous section are an important part of binaural simulation, but real-time binaural simulation encompasses more than just calculation of the binaural signals. This section will discuss different methods of
transmitting the binaural signals to the listener’s ears, the use of head tracking to create a real-time simulation, and interpolation between different HRTFs.

The exhaustive method described in the previous section calculates the binaural signals to represent a vibrating structure. These pre-calculated binaural signals can be directly transmitted to headphones to give a static listener a three-dimensional acoustic experience. However, simply playing the binaural signals over stereo speakers will not produce those binaural signals in the ears of the listener. The HRTF between the speakers and the listener must be accounted for in the binaural simulation, and the angle of the speakers must be optimized. The binaural signals are filtered with speaker filters, which cancel out the transfer function between the speakers and the listener. The optimized speaker configuration is termed the stereo-dipole method and will be presented in detail in Appendix 1 and headphone simulation will be further discussed in Appendix 3.

In addition to not accounting for the mechanism of transmitting the binaural signals to the listener’s ears, the binaural calculations mentioned in the previous section use only pre-calculated signals and a static listener. These calculations can be included in a more immersive simulation in which the listener is free to move and the listener’s movements affect the binaural calculations in real-time.

The schematic in figure 2.10 depicts a real-time simulation\textsuperscript{36,37} that is presented over headphones. As in the previous section, the code inputs are the monopole accelerations. The exhaustive method of figure 2.9 is performed for the current head position and angular orientation, denoted by $x_h$, $y_h$, $z_h$, $\theta_h$, $\phi_h$, and $\xi_h$. A head-tracking unit detects the listener’s position and continually feeds the current position into the computer. While the computer is calculating the sound due to the current head position, the headphones play the sound from the previous head tracker position. A real-time simulation using headphones was conducted by Knöfel and will be presented in Appendix 4.
It is important to note that the reduced binaural calculation methods (presented in chapter 3) can easily be substituted for the exhaustive method without altering the rest of the binaural simulation process. Also, if the stereo-dipole method is used rather than headphones, the simulation process will include filters to counteract the effects of cross-talk cancellation (see Appendix 1).

In addition, the measured HRTFs are not available for every point in a continuous field. In order to create a binaural signal at an angular position for which there is not a measured HRTF, one must interpolate between known HRTFs. Several researchers have investigated the error caused by interpolating HRTFs.\textsuperscript{19, 20, 38-41} In this work, the researchers interpolated the HRIRs to a $1^\circ$ resolution in the pre-processing stage. This degree of resolution was assumed to be sufficiently small because it is less than the smallest angle change that a human can detect. Bi-linear interpolation was implemented. That is, the HRIRs were linearly interpolated between azimuth angles and between elevation angles. Before interpolating, the interaural time delay (ITD) was removed (see Appendix 5) in order to reduce interpolation errors. After interpolation, the ITD is added back into the HRIRs.

\section*{2.6 Conclusion}

Each part of the binaural simulation process has been thoroughly discussed. A computational plate model was presented that determines the surface accelerations of a vibrating plate. Sound radiation from a monopole, group of monopole sources, and array of monopoles representing a vibrating structure was also introduced. In addition, the
application of HRTFs and the exhaustive method of binaural simulation were defined. The last section of this chapter presented the elements that combine to produce a real-time binaural simulation.
CHAPTER 3: REDUCTION METHODS

The exhaustive method of calculating the binaural signals (figure 2.9) models a structure as an array of vibrating monopoles, applies the HRTFs to each of these sources, and sums the resulting signals. Since thousands of monopoles are required to accurately represent the characteristics of a vibrating wall, the exhaustive method requires thousands of convolutions per iteration. For a sample rate of 44.1 kHz, each iteration must be carried out in less than 23 $\mu$s. Because it is currently impossible to perform thousands of convolutions in 23 $\mu$s, the objective of the author’s research has been to reduce the number of calculations required in order to determine the binaural signals of a vibrating structure. Three reduction methods will be discussed: Singular Value Decomposition (SVD), Wavenumber Filtering, and Equivalent Source Reduction (ESR).

3.1 Singular Value Decomposition

Singular value decomposition\(^{12,13,42}\) was the first reduction method to be investigated. By breaking down the matrix of HRTFs into three separate matrices, the convolution operation can be replaced by scalar multiplications. Since convolution is computationally expensive, a great deal of processing time can be avoided by a small amount of pre-processing. This discussion describes the theory behind SVD, the schematic of the SVD reduction method, and the number of singular values that are required in order to accurately reproduce the matrix of HRTFs.

3.1.1 Description of SVD reduction method

Singular value decomposition separates one matrix (in this case a matrix of HRTFs) into three separate matrices:

$$[H] = [U][\Sigma][V]^\top$$  \hspace{1cm} (3.1)

where $H$ is an $m$ by $n$ matrix of the HRTFs for every elevation and azimuth angle in one degree steps, $U$ is an $m$ by $m$ unitary matrix of left singular vectors, $\Sigma$ is an $m$ by $n$ diagonal matrix of singular values, and $V$ is an $n$ by $n$ unitary matrix containing right
singular vectors. While singular value decomposition can equivalently be performed in the frequency domain (i.e., on the HRTFs), the following analysis is in the time domain.

The original matrix of HRIRs for the left ear at 0° elevation is shown in figure 3.1. Notice that when the sound approaches the left ear directly (that is, when the angle between directly ahead and the approach of the sound is -90°), the sound is loud and contains high frequency content. When the sound comes from the other side of the head (90°), the sound level is lower (due to the IID), there is a time delay (ITD), and the signal contains mostly low frequencies because higher frequencies don’t wrap around the head.

![Figure 3.1: (a) Contour plot of the HRIRs for 0° - 360° azimuth, and (b) same HRIRs for only the 90° and -90° azimuth cases](image)

The first step taken in the SVD method was to create a function to describe the time delay of the HRIRs for each angle and remove that time delay from each impulse response in the matrix of HRIRs (figure 3.2). This time delay removal causes the matrix of HRIRs to be a smoother angular function so that the singular value decomposition can be performed with greater accuracy. Note that this operation is not necessary for HRIRs having a minimum phase representation. The time delay removed in this operation is added to the retarded time delay ($r/c$) in the radiation model. Appendix 5 provides an in-
depth discussion of the different methods that were used to model the time delay as a function of azimuth angle.

The singular value decomposition of the HRIR matrix (without time delays) is shown in figure 3.3. Notice that the left singular vector matrix, $U$, contains angular information for each singular value, $\sigma_i$. The matrix of singular values is simply a diagonal matrix of singular values that diminish in value as $i$ becomes larger. The time information for each singular value is contained in the right singular vector matrix. When these three matrices are multiplied together, the matrix of HRIRs (without time delays) will be the result. The advantage of SVD is that an approximate HRIR matrix can be obtained using only a few singular values, which is important when a large number of sources are analyzed.
3.1.2 Schematic of SVD reduction method

Recall the schematic of the exhaustive code in figure 2.9. Shown in figure 3.4 is a similar schematic of the SVD reduction method. Note that instead of simply convolving the HRIRs with the radiated pressure for each of the $N$ sources, the SVD method expands the application of the HRIRs into four faster operations that are performed for each of $M$ important singular values. The first operation is a scalar multiplication, in which the $N$ radiated pressures are multiplied by the corresponding element in the $N$ length vector $U_i$ (for $i = 1:M$). After applying the angle dependent left singular vectors, the signals for each $i^{th}$ singular value can be summed (before being filtered by the right singular vectors). This summing step is possible because the $V_i$ filters and singular values ($\sigma_i$) are not angle dependent. In addition, the scalar singular values, $\sigma_i$, can be pre-multiplied by the right singular vectors, $V_i$ (for $i = 1:M$), and the resulting filters can be convolved with the $i$ summed signals. The $i$ filtered signals can then be summed in order to produce the left ear signal. A similar process is carried out for the right ear signal.

![Figure 3.4: Schematic of the SVD Method](image-url)
The right and left singular vectors $U_i$ and $V_i$ are found in the $i^{th}$ columns of the matrices $U$ and $V$, respectively. While the left singular vectors change with observer angle, the right singular vectors, $V_i$ for $i = 1:M$, are independent of the observer location and angle. This is advantageous because the right singular vectors do not need to be found in a look-up table in real-time. In addition, these static filters ($V_i * \sigma$) could also be coded into hardware, which would significantly reduce the computation time required by the SVD method, simply because encoded hardware processes much faster than software. An additional significant computational advantage is found when $M$ is chosen such that $M < N$. When $M < N$, the number of convolutions required by the SVD method is much less than that required by the exhaustive method. Since the convolution process is the most computationally expensive part of the binaural simulation, a considerable reduction in computation time can be achieved.

3.1.3 Number of important singular values

Figure 3.5 shows the magnitude of the normalized singular values versus singular value number for four different frequency ranges. Notice that as the low-pass filter is applied to the HRTFs at successively lower frequencies, the number of important singular values also decreases.
Figure 3.5: Singular value number verses magnitude for four frequency ranges

The error coefficient, $\alpha$, is the log magnitude of the least-squares error associated with recreating the HRIR matrix using $M$ singular values:

$$\alpha = 10 \log_{10} \left( 1 - \frac{\sum_{i=1}^{M} \sigma_i^2}{\sum_{j=1}^{128} \sigma_j^2} \right)$$  \hspace{1cm} (3.5)

in which there are a total number of 128 singular values because the MIT HRIRs are 128 samples long. The number of important singular values, $M$, can be found by choosing a suitable value for $\alpha$ and plotting $M$ versus $\alpha$. Figure 3.6 shows $M$ versus $\alpha$ for the same frequency ranges as in figure 3.5. Selecting a criterion of $\alpha = -20$ dB, then requires the use of only three singular values for the 2500 Hz frequency range. Another criterion in selecting the number of singular values, $M$, is the correlation between the SVD method and the exhaustive method, which will be discussed in chapter 4.
3.2 Wavenumber Filtering

Wavenumber decomposition consists of transforming finite waves (such as on the surface of a finite plate) into the sum of infinite waves. This is analogous to transforming a finite signal in the time domain into the sum of infinite sine waves, each of a different frequency. In fact, the Fourier Transform is used for both. The only difference is that instead of transforming a signal from the time domain to the frequency domain, wavenumber decomposition transforms a signal from the spatial domain to the wavenumber domain. For an introduction to wavenumbers, see Appendix 6.

Figure 3.7 shows the two-dimensional wavenumber decomposition of a plate being driven at 550 Hz. (The measured data used to generate this plot will be described in section 4.1.) At this frequency, the acoustic wavenumber ($\frac{\omega}{c}$) is 10 rad/m and the bending wavenumber is 22 rad/m. As described in Appendix 6, only wavenumbers that are less than the acoustic wavenumber will radiate sound. These are designated as the
radiating wavenumber components. The theory behind wavenumber filtering is that since only some of the components radiate sound, for acoustic purposes it is only necessary to perform computations using these components.\textsuperscript{32, 33}

Figure 3.7: Wavenumber Decomposition of plate at 550 Hz

Wavenumber filtering (figure 3.8) is accomplished by first transforming velocity information from the spatial domain into the wavenumber domain via the Fourier transform. This signal is then low-pass filtered in the wavenumber domain (with the cutoff at the acoustic wavenumber) to isolate the radiating components. Next, the inverse Fourier transform is performed, converting the signal back into the spatial domain. Finally, the signal is re-sampled at a lower rate of spatial sampling. The end result is that the original plate measurements are replaced with a fewer number of sources, while still accurately reproducing the sound field.
There are two main problems with this approach. First, the resampling process causes apparent sources to appear outside the actual edges of the plate. This is visible in the bottom right plot in figure 3.8, in which sources of velocity are seen as far as 2 m from the center of the plate, which is outside the edges of the 1.4 m by 1.4 m plate. Second, wavenumber filtering can only be utilized for relatively simple geometries, such as a flat plate. For preliminary analysis, that is not an issue, but for more complex vibrating structures, such as the walls of the space station, wavenumber filtering will not be applicable. Since the future applications of this research include more complicated structures, wavenumber filtering will not be used. However, this is essentially what the equivalent source reduction method (section 3.3) will achieve.

### 3.3 Equivalent Source Reduction

Similar to wavenumber filtering, equivalent source reduction (ESR) replaces a vibrating structure with a small number of equivalent sources, which accurately reproduce the sound field. This reduction is accomplished by calculating the velocity profile across an imaginary evaluation surface due to the original sources and matching that velocity profile with one calculated from a reduced set of equivalent sources. In
addition, the accelerations of the reduced set of equivalent sources are calculated in the
pre-processing stage (rather than in real-time), so the ESR method offers a pure reduction
in the amount of computation time.

Another advantage of the ESR method is that it can be applied to any general
shaped structure and can be used to replace sources that are radiating sound into either
the free field or an enclosure. The generalized calculations involved in ESR will be
described first, followed by several of the cases to which ESR can be applied. The case
in which ESR is applied to a planar structure will be validated with measured data in the
following chapter. Much of the credit for the development and implementation of the
ESR method is owed to Marty Johnson\textsuperscript{43,44}.

3.3.1 Description of Equivalent Source Reduction

The equivalent source technique replaces a vibrating structure with a reduced set
of equivalent sources, which produce an equivalent sound field. This is achieved by
creating an evaluation surface between the source and the listener (as in figures 3.9 –
3.11). The complex accelerations of the vibrating structure have been measured or
calculated a priori. Knowing these source accelerations, one can determine the complex
velocities normal to the evaluation surface at several evaluation points. The velocity
profile on the evaluation surface due to a reduced set of equivalent sources can be
similarly determined. These equivalent source accelerations are chosen so that the two
velocity profiles match in a least squares sense. Since the calculations are carried out in
the pre-processing stage (rather than in real-time), the equivalent source technique can be
performed in either the time or frequency domains. The frequency domain version was
implemented and is shown here because it is simpler than the analogous solution in the
time domain.

Consider a general vibrating structure that is radiating sound into the free field.
The complex accelerations of each of these monopoles (at one frequency, $\omega$) are all
contained within an $N$ length complex vector $a(\omega)$. The complex velocities at the $L$
evaluation points along the evaluation surface, $S_E$, enclosing the group of sources are
contained within the $L$ length complex vector, $u(\omega)$, defined as:

$$u(\omega) = T(\omega)a(\omega)$$  \hspace{1cm} (3.9)
where \( \mathbf{T}(\omega) \) is an \( L \) by \( N \) matrix of complex transfer functions between the acceleration of each source and the normal velocity at each evaluation position.

The evaluation velocities due to a reduced set, \( N_E \), of equivalent sources can be similarly found:

\[
\mathbf{u}_E(\omega) = \mathbf{T}_E(\omega)\mathbf{a}_E(\omega)
\]

(3.10)

where \( \mathbf{u}_E(\omega) \) is an \( L \) length complex vector of velocities normal to the evaluation surface, \( \mathbf{T}_E(\omega) \) is an \( L \) by \( N_E \) matrix of complex transfer functions between the acceleration of each equivalent source and the evaluation velocities, and \( \mathbf{a}_E(\omega) \) is an \( N_E \) length complex vector containing the accelerations of each equivalent source, all of which are defined at the single frequency, \( \omega \).

As long as the accelerations \( \mathbf{a}_E(\omega) \) of equivalent sources are chosen such that the particle velocities \( \mathbf{v}_E(\omega) \) are the same (or very similar) to the velocities \( \mathbf{v}(\omega) \), then the acoustic field outside the surface \( S_E \) will be the same as the original field. This can be deduced from the Kirchoff-Helmholtz equation\(^{25, 31, 32} \) assuming that there are no other sources acting in the field. In order to find the acceleration \( \mathbf{a}_E(\omega) \) that minimizes the difference between the true velocity and the simulated velocity, the following error minimization method will be used.

\[
\mathbf{a}_E(\omega) = \left[ \mathbf{T}^H_E(\omega)\mathbf{T}_E(\omega) \right]^{-1}\mathbf{T}^H_E(\omega)\mathbf{T}(\omega)\mathbf{a}(\omega)
\]

(3.11)

This method minimizes the difference (in terms of least squares) between the true and simulated velocities.

If the transfer function terms in the above equation are combined such that:

\[
\mathbf{C}(\omega) = \left[ \mathbf{T}^H_E(\omega)\mathbf{T}_E(\omega) \right]^{-1}\mathbf{T}^H_E(\omega)\mathbf{T}(\omega)
\]

(3.12)

the end result is an \( N \) by \( N_E \) matrix \( \mathbf{C}(\omega) \) that converts the \( N \) original source accelerations into \( N_E \) equivalent source accelerations for the single frequency \( \omega \) as in:

\[
\mathbf{a}_E(\omega) = \mathbf{C}(\omega)\mathbf{a}(\omega)
\]

(3.13)

This conversion matrix, \( \mathbf{C}(\omega) \), can be found for each frequency in the frequency range of interest until an \( N \) by \( N_E \) matrix of frequency domain filters is obtained. The conversion matrix \( \mathbf{C}(\omega) \) can then be transformed into the time domain using an inverse Fourier
transform (to create $C(t)$). Since this conversion is done in a pre-processing stage the
time domain filters contained in $C(t)$ are not constrained to be causal.

This matrix of time domain filters, $C(t)$, can then be used to convert the time
domain accelerations $a(t)$ into a reduced set of $N_E$ equivalent source accelerations $a_E(t)$.
These equivalent source accelerations can be directly used in the sound radiation
equations 2.8 to 2.12. In addition, these equivalent sources can be directly substituted for
the original inputs into the exhaustive method of figure 2.9. Since the number of sources
is reduced, the number of overall convolutions is reduced, thereby significantly reducing
the number of computations required in the ESR method as compared with the exhaustive
method.

3.3.2 ESR for a planar surface

Consider a vibrating planar structure (figure 3.9) that is set in an infinite baffle
and radiates sound into the free field. A flattened hemispherical evaluation surface, $S_E$,
can be created to surround the vibrating panel. Every equation in the ESR process
described in the previous section applies to the case of the vibrating planar structure.
However, the calculation of the transfer function matrix, $T(\omega)$, is case dependent, and
will be discussed here for the specific case of a planar structure.

![Diagram of Equivalent Source Reduction](image)

Figure 3.9: Equivalent source reduction applied to a vibrating panel

The elements of the transfer function matrix, $T(\omega)$, can be found by calculating
the particle velocities across the evaluation surface and taking the transfer functions
between these velocities and the known plate accelerations. Because the plate can be
modeled as a group of vibrating hemispheres, the pressure field normal to the evaluation surface can be found by considering the sound radiation from each monopole to each evaluation point. Equation 2.9, rewritten here, dictates that the sound radiated from a source positioned at $\mathbf{x}_j$ will create pressure $p_{ij}$ at an evaluation position $\mathbf{x}_i$ given by:

$$p_{ij}(\omega) = a(\omega) \left( \frac{\rho_0 S}{2\pi r} \right) e^{i(\omega t - kr)}$$

where

$$r = |\mathbf{x}_i - \mathbf{x}_j|$$

The pressure $p_{ij}$ can be used to find the normal particle velocity $v_{ij}$ at an evaluation position $\mathbf{x}_i$ by employing the fluid momentum equation in the direction normal to the evaluation surface:

$$v_{ij}(\omega) = \left( -\frac{1}{j\omega \rho_0} \right) \frac{\partial p_{ij}(\omega)}{\partial n_i}$$

where $n_i$ is the vector normal to the surface $S_E$ at the evaluation position.

Calculation of the matrix of transfer functions $T(\omega)$ is then a simple matter of taking the transfer functions between the plate accelerations $a(\omega)$ and each normal particle velocity $v_{ij}$. The matrix of transfer functions $T_E(\omega)$ is calculated by applying the same equations to a reduced set of equivalent sources.

### 3.3.3 ESR applied to complex sources

In addition to reducing the number of sources required to represent a vibrating structure, equivalent source reduction can be applied to more complex geometries than the exhaustive method. In the case of a planar structure, the exhaustive method replaces the plate with an array of vibrating hemispheres. However, for a complex-shaped structure (such as the one in figure 3.10), the structure would interfere with the sound field produced by the monopoles. The equivalent source technique is able to replace the original structure with a group of monopoles in the free-field, which essentially assumes that the structure is acoustically invisible.
The only difficulty in applying the ESR method to the case of a complex shape is in the calculation of the velocity across the evaluation surface due to the original sources (the velocity across the evaluation surface due to the equivalent sources is trivial and can be accomplished with the free field equations, 3.14 – 3.16). Since, as mentioned before, the original sources cannot be replaced with free field monopoles, another method must be used to find the evaluation surface velocity. In this case, several different methods can be applied in order to find the velocity normal to the evaluation surface: finite element methods, boundary element methods, or another application of the equivalent source technique.25

This additional equivalent source application would consist of creating a secondary evaluation surface on the surface of the complex structure. The velocity across this evaluation surface is already known (measured data) and can be matched with the velocity due to a secondary set of equivalent sources, which can be positioned at any arbitrary positions inside or on the evaluation surface. These secondary equivalent sources are free-field monopoles, so the velocity on the primary evaluation surface (at a distance from the original structure) can easily be calculated. Note that the secondary equivalent sources are not a reduced set of sources. The source reduction is
accomplished by placing the primary evaluation surface at a distance from the original surface. The air gap acts as a natural low-pass wavenumber filter, allowing only the acoustically important components to be represented by the reduced set of primary equivalent sources.

ESR can also be applied to a structure that is radiating sound inward into an enclosure (figure 3.11)\textsuperscript{33, 45, 46}. That is, the surface of the structure is vibrating and creating sound which is reflected off the other inner surfaces of the structure and detected by an observer positioned inside the vibrating enclosure. Similar to the case of a general surface radiating sound into the free field, the surface cannot be modeled as a group of monopoles by assuming the measured velocities at several nodes on the structure are equal to the velocities of monopoles positioned at the same nodes. In addition, the structure is not radiating sound into a free field, so the sound reflections must be accounted for.

![Diagram of Equivalent Source Reduction](image)

Figure 3.11: Equivalent source reduction applied to a vibrating enclosure

Boundary element methods and finite element analysis can again be used to find the velocity across the evaluation surface due to the direct sound radiation from the
vibrating walls. In addition, the sound reflections can be modeled using image sources (see Appendix 2) and the velocities across the evaluation surface due to the reflected sound can be determined using equation 3.14. Because the system is linear, the velocities across the evaluation surface due to direct radiation can be added to the velocity due to reflection. The result of this linear superposition is the total velocity on the evaluation surface due to direct and reflected sound. Equation 16 can then be employed to determine the particle velocity normal to the evaluation surface, which results in the transfer function matrix, \( T(\omega) \). As in the case of a general structure radiating into the free field, the matrix \( T_\text{E}(\omega) \) can be calculated using the free field equations (3.14 – 3.16).

Using this method the reflective nature of the enclosure is accounted for completely in a pre-processing stage and replaced by sources radiating sound in a free field (i.e., no image sources are required).

3.4 Conclusion

Three different reduction methods have been presented: the SVD method, wavenumber filtering, and the ESR method. Since wavenumber filtering causes imaginary sources to appear outside the edges of the panel and cannot be used for complex sources, it was deemed unsuitable for this research. However, the ESR method, which reduces the number of sources effectively by filtering out the non-radiating wavenumber components, is able to be used for complex sources and does not create sources outside the vibrating structure. In addition to the pre-processing ESR method, the real-time SVD method is a promising method for reducing the number of real-time computations in binaural simulation. In addition to the exhaustive method, both the SVD and ESR reduction methods were implemented using measured plate data, which will be described in the following chapter.
CHAPTER 4: IMPLEMENTATION USING MEASURED DATA

Having developed the theory behind the exhaustive, SVD, and ESR methods of calculating the binaural signals, it remains to experimentally verify the accuracy and effectiveness of each method. This has been accomplished by implementing each method with plate accelerations that were measured in a transmission loss facility. The results found by implementing the exhaustive method with measured plate data will be compared with measured binaural signals. In addition, the binaural signals found through the reduced methods will be compared with the exhaustive method results. Auditory tests were also conducted with the stereo-dipole method in order to compare the accuracy of each binaural calculation method.

4.1 Description of Measured Data

Grosveld collected the measurements used to implement the exhaustive and reduced methods at the NASA Langley Research Center’s transmission loss (TL) facility. The consistency of these measurements was verified by graphically comparing the pressure recorded at a reference microphone during two separate tests. Finally, the frequency range of accuracy of the measured data was confirmed by performing wavenumber decomposition on the measured plate data.

4.1.1 Measurement Collection

Three types of measurements were collected in the NASA Langley TL facility, which are structural, acoustic, and binaural transfer functions. An aluminum panel was mounted in the window of the transmission loss facility and subjected to mechanical excitation. The structural transfer functions were taken between the force input and the accelerations of 529 nodes on the panel, which were measured with a scanning laser vibrometer. In addition, a microphone was positioned inside the anechoic side of the transmission loss facility to measure the acoustic transfer functions between the input force and the free field pressure at various locations. Finally, the binaural transfer functions were taken between the force input and the binaural signals measured at
microphones inside the ears of a Knowles Electronic Manikin for Acoustic Research (KEMAR).

These measurements were all collected in the NASA Langley Research Center Structural Acoustic Loads and Transmission (SALT) facility. The TL facility consists of a 337 m³ reverberant chamber and a 278 m³ anechoic chamber joined by a TL window which accommodates 1.415 m by 1.415 m test structures. The anechoic chamber side of the TL facility is acoustically dampened by over 4850 open-cell, polyether-polyurethane acoustic wedges that cover the walls, ceiling and floor. A 4.9 mm thick aluminum test panel was secured in the TL window using four 0.5 in. thick steel bars and clamped to the TL window with 56 evenly spaced bolts. In addition, four reflective ball bearings mounted on the window frame were used as reference points and a trilateration procedure was used to calculate distances within an accuracy of four millimeters. A schematic of the transmission loss facility is shown in figure 4.1. A more detailed description of the SALT facility has been provided by Grosveld.

Figure 4.1: Schematic of the transmission loss facility
A B&K Mini Shaker Type 4810 provided the mechanical excitation. All of the structural, acoustic, and binaural transfer functions were found for two separate force input locations. The first force location was at the center of the plate, designated (0,0,0) in Cartesian coordinates. An off-center location was also used at (-502 mm, -500 mm, 0 mm). The coordinate system and panel are shown in figure 4.2 and the orientation of the shaker is shown in figure 4.3.

Figure 4.2: Coordinate system used for data measurements

Figure 4.3: Mechanical excitation on the reverberant side of the panel
In addition to the mechanical shaker, four accelerometers (figure 4.4) were positioned on the reverberant side of the panel. One of the accelerometers was fixed to the frame (in order to detect flanking), one was connected to the mechanical shaker, and the other two were placed in generic locations on the panel. Since measurements were taken for two shaker locations, five accelerometer positions are listed in table 4.1.

![Figure 4.4: Three of the accelerometers attached to the aluminum panel](image)

Table 4.1: Coordinates of measurement locations on aluminum test panel

<table>
<thead>
<tr>
<th>Location</th>
<th>Transducer</th>
<th>X-coordinate (mm)</th>
<th>Y-coordinate (mm)</th>
<th>Z-coordinate (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location 1</td>
<td>force/accelerometer</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Location 2</td>
<td>accelerometer</td>
<td>-118</td>
<td>118</td>
<td>0</td>
</tr>
<tr>
<td>Location 3</td>
<td>accelerometer</td>
<td>-353</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Location 4</td>
<td>accelerometer</td>
<td>-712</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Location 5</td>
<td>force/accelerometer</td>
<td>-502</td>
<td>-500</td>
<td>0</td>
</tr>
</tbody>
</table>

In addition to measuring the plate acceleration at a few points with accelerometers, a scanning laser vibrometer (figure 4.5) was used to detect the plate acceleration at 529 points on the plate. This tight grid was chosen so that the measurement spacing would be less than half the structural wavelength up to 2000 Hz. The scanning laser vibrometer was chosen instead of a 23 by 23 grid of accelerometers because it is a non-invasive instrument (it does not affect the plate response). This frequency range was chosen so that both IID and ITD cues could be investigated. (The primary directional cue up to 1000 Hz is ITD, after which IID becomes more important.)
The following plot (figure 4.6) is the sum of the squares of the plate velocities measured with the scanning laser vibrometer, assuming a white noise force input. Notice that the frequency response has very sharp resonance peaks. This is because the test plate was very lightly damped. In fact, this plot actually underestimates the damping because exponential windowing was added to the vibrometer measurements.⁵
In addition to measuring the structural vibration of the plate, acoustic transfer functions were also collected. A B&K free-field type 4165 microphone was placed at locations 1-4 listed in table 4.2 and transfer functions, auto spectra, and coherence functions were collected. The resulting transfer function from the input force to a microphone is shown in figure 4.7a at the position shown in figure 4.7b (Location 1). The low damping level of the test plate noticed from the structural measurements is also visible in this acoustic transfer function. In addition to the free field microphone, a reference microphone (B&K type 4134) was positioned at location 5 and remained there throughout the measurement collection.

Table 4.2: Free field and reference microphone positions

<table>
<thead>
<tr>
<th>Location</th>
<th>X-coordinate (mm)</th>
<th>Y-coordinate (mm)</th>
<th>Z-coordinate (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location 1</td>
<td>0</td>
<td>0</td>
<td>925</td>
</tr>
<tr>
<td>Location 2</td>
<td>-1234</td>
<td>0</td>
<td>1953</td>
</tr>
<tr>
<td>Location 3</td>
<td>0</td>
<td>0</td>
<td>3048</td>
</tr>
<tr>
<td>Location 4</td>
<td>0</td>
<td>0</td>
<td>3229</td>
</tr>
<tr>
<td>Location 5</td>
<td>-1526</td>
<td>524</td>
<td>1147</td>
</tr>
</tbody>
</table>

![Figure 4.7](image.png)

Figure 4.7: (a) Transfer function between the input force and the microphone in the position shown in (b)

The third type of measurements that were collected is the acoustic transfer functions between the input force and pressure microphones in the left and right ears of a KEMAR dummy head (figure 1.2). The KEMAR manikin was positioned in several different angles and positions with respect to the panel in order to study different angular
effects. This was made possible by mounting the manikin on a support structure (figure 4.8) that allowed rotation in the \( x \), \( y \), and \( z \) directions.

![Figure 4.8: KEMAR mounted in support structure](image)

As with the free field microphone measurements, auto spectra, coherence functions, and transfer functions were measured for a pseudo-random shaker input. In addition, time histories were acquired for pulse and broadband random inputs to the shaker. The transfer function from the input force (at the center of the plate) to the left ear of the KEMAR manikin rotated in the horizontal direction +30° (according to the coordinates in figure 4.1) is shown in figure 4.9a (with the orientation of figure 4.9b).

![Figure 4.9: (a) Transfer function from input force to left ear, configured as in (b)](image)
4.1.2 Measurement Consistency

Since the measured data described in the previous section is used to validate the exhaustive and reduced methods of calculating the binaural signals (sections 4.2 and 4.3), the consistency and accuracy of the data is important. The following plot (figure 4.10a) shows the transfer function from the input force to a reference microphone (positioned at location 5 and according to the configuration in figure 4.10b) during two separate tests (numbers 22 and 30, as designated by Grosveld5). Note that the position of the reference microphone was not changed at all between tests. The primary difference between the two measurements was the height of the resonance peaks, which corresponds to different estimates of the plate damping. This could be due to slight shifts in natural frequencies between measurements. If this is the actual cause of the resonance peak height differences, it implies that more frequency bins would be needed to accurately determine the height of these sharp resonance peaks, which in the time domain corresponds to longer measurement collection windows. There are also some discrepancies at low dB, but these are expected because there is a low signal to noise ratio at these frequencies.

![Transfer Function Plot](image)

Figure 4.10: (a) Consistency between two reference microphone measurements and (b) reference microphone orientation
In addition to determining the consistency of the transfer function measurements, the accuracy of the tests was also investigated. The overall coherence of each of the tests was very high. Figure 4.11 shows the coherence of the transfer function from the input force to the reference microphone (same configuration as in figure 4.10b). Notice that there are several points in which the coherence is rather low. These points correspond to low input levels and therefore do not decrease the accuracy of the measured data.

![Figure 4.11: Coherence of the transfer function from the input force to the reference microphone](image)

**4.1.3 Wavenumber Decomposition**

After verifying the consistency of measurements, the researchers used wavenumber decomposition\(^\text{32,33}\) to verify the frequency range for which the data is accurate. The plate was measured every 0.06m (figure 4.12), which corresponds to a sample wavenumber of \(k_s = \frac{2\pi}{\lambda} = 102\) rad/m. Spatial aliasing will occur if the wavenumber on the plate is greater than half the sampling wavenumber. Figure 4.13 shows the wavenumber decomposition of the plate velocities at 550 Hz. At this frequency, the bending wavenumber of the plate can be estimated to be 16 rad/m, where \(k_b\) is defined according to equation 4.1:
\[ k_b = \sqrt{k_{bx}^2 + k_{by}^2} = \left[ \frac{\omega^2 m}{D} \right]^{\frac{1}{4}} \] (4.1)

where \( m \) is the mass of the plate, defined as \( \rho L_x L_y \). The bending stiffness of the plate, \( D \), is defined as:

\[ D = \frac{E h^3}{12(1 - \nu^2)} \] (4.2)

where \( E \) is the modulus of elasticity of the plate, \( h \) is the plate thickness, and \( \nu \) is the Poisson’s ratio of the plate. The plate properties used to calculate the bending stiffness are shown in table 4.3.

Figure 4.12: Configuration of the aluminum test panel

Table 4.3: Test plate properties

<table>
<thead>
<tr>
<th>Plate Parameter</th>
<th>Numerical Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modulus of elasticity, ( E )</td>
<td>( 7.1 \times 10^{10} ) N/m²</td>
</tr>
<tr>
<td>Plate density, ( \rho_s )</td>
<td>2700 kg/m³</td>
</tr>
<tr>
<td>Plate dimension in ( x ) direction, ( L_x )</td>
<td>1.41 m</td>
</tr>
<tr>
<td>Plate dimension in ( y ) direction, ( L_y )</td>
<td>1.41 m</td>
</tr>
<tr>
<td>Plate dimension in ( z ) direction, ( h )</td>
<td>0.0127 m</td>
</tr>
<tr>
<td>Poisson’s ratio, ( \nu )</td>
<td>0.33</td>
</tr>
</tbody>
</table>
Figure 4.13: Wavenumber decomposition of the plate accelerations at 550 Hz

Notice that the acoustic wavenumber, \( k_a = \frac{\omega}{c} \), is only 10 rad/m, which is smaller than the structure’s bending wavenumber. As stated in the discussion of wavenumbers in Appendix 6, only waves whose wavenumbers are less than the acoustic wavenumber will radiate sound to the far field. Figures 4.14 and 4.15 show the wavenumber decomposition of the plate at 2000 and 4000 Hz respectively. In figure 4.15 spatial aliasing occurs, in which the modal wavenumbers are “reflected” at the edges of the measurement sampling wavenumber. At 2000 Hz (figure 4.14) the bending wavenumber is equal to half the structural wavenumber and free of aliasing, proving that the data is accurate at frequencies up to 2000 Hz.
Figure 4.14: Wavenumber decomposition at 2000 Hz (no aliasing)

Figure 4.15: Wavenumber decomposition at 4000 Hz (aliasing visible)
4.2 Validation of Exhaustive Method

The goal of this research is to develop and evaluate the effectiveness of methods to reduce the number and complexity of binaural calculations. In order to evaluate the effectiveness of these reduction methods, they will be compared with the exhaustive method in section 4.3. First, the exhaustive method needs to be established as an accurate benchmark. To accomplish this, measured plate accelerations will be used as input into the exhaustive code (figure 2.9) and the resulting radiated pressure will be compared with the acoustic transfer function measured with the free field microphone. In a similar way, the binaural signals calculated through the exhaustive method (with measured plate data used as the input) will be compared with the binaural recordings measured at the right and left ears of the KEMAR. In this way, every step of the exhaustive method will be validated.

4.2.1 Radiation Model

A direct comparison (figure 4.16) can be drawn between the calculated and measured transfer functions from the input force at the plate center to a microphone at location 1 (as in figure 4.7b). The calculated transfer function was found through the exhaustive method (with measured plate velocities used as the inputs), and the measured transfer function was found from the measured pressure at the free-field microphone. So, the only difference between these two signals is that the radiation from the plate to the microphone was calculated using a computer program in one instance, and measured in the other. This will test the accuracy of the free field equations, which are equations 2.11 – 2.14.
Figure 4.16: Frequency domain comparison of measured and calculated radiated pressures

The measured and calculated transfer functions in figure 4.16 are satisfactorily similar. Notice, however, that there are two basic types of inconsistencies seen in this comparison: there are differences in the heights of the peaks, and there are discrepancies at low pressure levels. The differences in peak heights signify a difference in the amount of damping measured by the scanning vibrometer and the microphone. As mentioned previously, this is due to the exponential windowing that was applied only to the vibrometer measurements and to the extremely low damping on the plate, which requires a long measurement window in order to measure precisely.

There are also some discrepancies at low dB, but these are natural occurrences due to low system input level. As mentioned in the discussion of the theory behind the exhaustive method, the radiation model simply adds a time delay and gain to the signal. It is important to notice from this comparison that the magnitude of the measured and calculated frequency responses are the same. Looking at a comparison of the measured and calculated pressures in the time domain will allow a check of the time delay.
The following (figure 4.17) is a comparison of the calculated and measured impulse responses from the input force at the plate center to a microphone about 1m from the plate (at location 1, as in the frequency domain comparison of figure 4.16). Both of the impulse responses are band-pass filtered between 40 and 2000 Hz in order to represent only the frequencies for which the measured data is accurate. While this view of the impulse response is too broad to see the initial time delay, it allows examination of the end of the impulse response. Notice that the measured response rings longer than the pressure calculated using the exhaustive method. This is expected because of the differences in damping level estimated by the two transfer function measurements. As stated in the frequency domain comparison, the low damping level of the plate requires a longer measurement window (longer than 1.6 seconds) in order to accurately measure the damping.

![Figure 4.17: Time domain comparison of the measured and calculated radiated pressures (extended view)](image)

A closer view of the initial part of this impulse response (figure 4.18) shows high correlation (71% correlation) between the time delay of the impulse responses. This is
significant because any errors in the radiation model would correspond to differences between the time delay and gain of the calculated and measured responses. The closely matching time delay (as visible through the impulse response comparison) and magnitude (as visible through the transfer function comparison) verify that the radiation model part of the exhaustive method is correct.

![Time domain comparison of the measured and calculated radiated pressures](image)

Figure 4.18: Time domain comparison of the measured and calculated radiated pressures (close-up view)

Were these measurements to be repeated, the low damping level of the plate would have to be taken into consideration. As mentioned previously, a longer time window would allow a more accurate and consistent measurement of the plate damping. In addition, more damping could be added to the plate by securing cross-bars to the plate or increasing the plate thickness. In any case, the accuracy and consistency of this data is sufficient to verify the exhaustive and reduced methods.

### 4.2.2 HRTF Calculations

In order to check the HRTF application part of the exhaustive method, the researchers compared measured and calculated transfer functions between the input force
and the binaural signals at the ears of a dummy head. The calculated binaural signals were found through the exhaustive method using the plate accelerations that were measured with the laser vibrometer as the input to the exhaustive code of figure 2.9 (as with the radiation model check). The following discussion details how these calculated binaural signals compare with signals that were actually measured at the ears of a KEMAR dummy head. For the following analysis, the head and plate were assumed to be configured according to figure 4.19, in which the force is at the plate center and the KEMAR head is rotated $+30^\circ$ in the horizontal, and $0^\circ$ in the vertical plane (as per the coordinate system in figure 4.2). In addition, the KEMAR head is situated at location 1 (as described in table 4.2) and the following transfer function and impulse response are between the input force and the pressure at the KEMAR’s left ear.

![Figure 4.19: Head and plate configuration for HRTF application check](image)

Figure 4.20 shows the frequency domain comparison between the measured and calculated transfer functions from the input force to the left ear of the KEMAR head. As before, the calculated transfer function refers to the transfer function gleaned from the exhaustive method, with the measured plate accelerations used as inputs into the exhaustive code, while the measured transfer function was found from the actual measured pressure at the KEMAR’s left ear (in the configuration of figure 4.19). As in the radiation model check of the previous section, the frequency domain comparison allows a check of the overall magnitude of the transfer function. Although figure 4.20
shows the same discrepancies at low dB and the same problems estimating the low damping levels of the plate (as visible in the sharp resonance peaks), the magnitudes of the measured and calculated transfer functions match very well.

Figure 4.20: Frequency domain comparison of measured and calculated left binaural signal

Figure 4.21 shows the measured signal at the left ear of the KEMAR head compared with the left binaural signal calculated in the exhaustive method. As with the verification of the radiation model, this time domain comparison shows a good correlation (53% correlation) between the time delay of the measured and calculated impulse responses between the input force and the left ear (in the configuration of figure 4.19). Based on the magnitude correlation found in the frequency domain comparison and the time delay match found in the time domain, the application of HRTFs in the exhaustive method is verified to accurately match the measured data and is therefore an acceptable benchmark against which the reduction methods can be compared.
4.3 Validation of Reduction Methods

Having shown in section 4.2 that the exhaustive method accurately produces the binaural signals needed for binaural simulation (given the limits on accuracy of damping levels in the measured data), the reduced methods can be evaluated by comparing them with the exhaustive method. The measured plate accelerations will be used as inputs into both the exhaustive method (figure 2.9) and the SVD method (figure 3.4) and the resulting binaural signals will be compared. To assess the accuracy of the ESR method, the binaural signals resulting from the ESR and exhaustive methods will also be compared. As with the SVD method, the input into the exhaustive method will be the measured plate accelerations, and the input into the ESR method will be the equivalent source accelerations \( a_E(\omega) \) of equation 3.13, where \( a(\omega) \) are the measured plate accelerations. The two reduced methods will also be combined and the resulting binaural signals will be compared with those calculated through the exhaustive method.
Finally, the computation time associated with each method will be estimated, thereby assessing the computational effectiveness of the reduction methods. In this manner, both the accuracy and effectiveness of the reduction methods will be addressed.

4.3.1 SVD Method

In order to compare the exhaustive method of figure 2.9 and the SVD method of figure 3.4, both methods were implemented using the 529 plate accelerations collected with the scanning laser vibrometer. The head and plate were configured according to figure 4.22, that is, the force input was at the plate center and the head was rotated -30° (according to the convention of figure 3.1) and positioned at location 1. Although any number of singular values can be used in the SVD method implementation, the accuracy increases as more singular values are used (see figure 3.6 for error versus number of summed singular values). In the following comparisons, three singular values are used in the SVD implementation so that an acceptable error level can be maintained (-20 dB for a frequency range up to 2500 Hz) while minimizing the number of computations required.

![Diagram of head and plate configuration](image)

Figure 4.22: Head and plate configuration for reduction method validations

Figure 4.23 compares the left binaural signal as calculated through the exhaustive and SVD methods. As previously mentioned, only three singular values were used to reconstruct the matrix of HRTFs and the methods were compared over a frequency range of 0 Hz to 2000 Hz. Note that there is very good agreement between the two and that any
discrepancies occur at low dB. Also notice that the SVD method (with three singular values) is a closer match to the exhaustive method at lower frequencies. This low frequency matching occurs because the lower order singular vectors contain mostly low frequency information. More singular values are needed (figure 3.6) in order to accurately represent the higher frequencies. Since structural responses are typically represented by low frequency data, especially if numerical methods such as FEM are used to generate the data, three singular values will generally be enough for accurate reproduction.

![Graph of frequency domain comparison](image)

**Figure 4.23:** Frequency domain comparison of the left binaural signal (exhaustive and SVD methods)

The time domain comparison analogous to that of figure 4.23 is the impulse response from the input force to the left ear, shown in figure 4.24. Since the collected data is accurate only up to 2000 Hz (as shown in section 4.1.3), a low-pass filter at 1600 Hz was applied. The correlation between the SVD and exhaustive methods was 97.7%, while using only three singular values. From this result the authors concluded that
singular value decomposition is an accurate reduction method and can be effectively used in this research.

![Time domain comparison of the left binaural signal (exhaustive and SVD methods)](image)

**Figure 4.24**: Time domain comparison of the left binaural signal (exhaustive and SVD methods)

### 4.3.2 ESR Method

Having determined that the SVD method accurately reproduces the binaural signals, the ESR method was implemented and evaluated. In addition to comparing the binaural signals of the ESR and exhaustive methods, the pressures at the center of the head were also compared. This comparison was not carried out for the SVD method, since by definition (see figures 2.9 and 3.4) the radiated pressures are identical. Note that the equivalent source schematic is the same as the exhaustive method of figure 2.9, except that the number of input sources, \( N \), is reduced and equivalent source accelerations are used rather than measured ones. The equivalent accelerations, \( a_E(\omega) \), are obtained from equation 3.13, where \( a(\omega) \) are the measured plate accelerations.

The equivalent source method was implemented using an evenly spaced grid of equivalent sources to represent the radiation from the plate. The evaluation surface
(similar to that of figure 3.9) was a flattened hemisphere where the radius in the plane of the plate was 1.2m, and the radius normal to the plate was 0.6m. This surface was covered by 884 evenly spaced evaluation points. Although any number of equivalent sources can be chosen, this discussion focuses on two cases: a 4 by 4 grid and an 8 by 8 grid of equivalent sources. In addition, the pressure at the head center is compared with the exhaustive method for observer positions in both the near-field and the far-field.

A far-field comparison of the pressure spectra at a microphone in space is shown in figure 4.25. Notice that the 8 by 8 grid of replacement sources matches the original 23 by 23 grid of sources at frequencies up to 980 Hz. This is expected because the source spacing (0.176 m) is equivalent to half an acoustic wavelength ($\pi c/\omega$) at 980 Hz.

![Position of observer [0.6, 0.6, 0.4] 1m from center](image)

Figure 4.25: Pressure at far-field observer calculated through exhaustive and ESR methods

Notice that in the near-field (figure 4.26) the equivalent source method is not as accurate. The two pressure signals would diverge even further when HRTFs are applied in order to calculate the binaural signals. In addition, HRTFs vary greatly with distance in the near-field (less than 1 m from the head). Since the MIT HRTFs are measured in
the far-field, additional error will be introduced if these HRTFs are used to represent a virtual source in the near-field. Because of these difficulties and the fact that most applications for this research are in the far-field, the near-field will not be investigated further.

Figure 4.26: Pressure at close observer found through exhaustive and ESR methods

When only 16 equivalent sources are used (figure 4.27), the pressure observed in the far-field is accurate up to 490 Hz. Again, this is because the source spacing is equal to half the acoustic wavelength at 490 Hz.
Figure 4.27: Pressure at observer found by exhaustive and 16 source ESR methods

The following figure (4.28) shows the frequency domain comparison of the pressure at the left ear of a listener (configuration in figure 4.22). As with the radiated pressure comparison, the binaural signals calculated with the ESR method (using 64 equivalent sources) is very close to the exhaustive method up to about 1000 Hz.
Figure 4.28: Left binaural signal in the frequency domain (exhaustive method and ESR method with 64 sources)

Figure 4.29 depicts the time domain version of figure 4.28 (low-pass filtered at 800 Hz). Cross-correlation of these two signals yields an excellent correlation of 98.4%.
When only 16 sources are used, the frequency domain version of the left binaural signal (configuration in figure 4.22) that was calculated using the ESR method is accurate up to 490 Hz (figures 4.30 and 4.31).
In the time domain (with a low-pass filter at 400 Hz), the correlation is found to be 98%.
In summary, the equivalent source reduction method very accurately reproduces the binaural signals necessary for binaural simulation, but only up to the frequency that corresponds to an acoustic wavelength that is twice the source spacing. So, reducing the number of equivalent sources lowers the frequency range of accuracy, but maintains the accuracy over that smaller frequency range.

However, the equivalent source method can be implemented without decreasing the frequency range of accuracy. For example, in order to maintain accuracy up to 2000 Hz (the same accuracy of the measured data), the source spacing should be equivalent to half an acoustic wavelength ($\frac{\pi c}{\omega}$), which is 0.086 m. This measurement spacing can be achieved by replacing the 23 by 23 grid of original sources with a 16 by 16 grid of equivalent sources. So the 529 original sources can be reduced to 256 equivalent sources while maintaining the same frequency range of accuracy, a source reduction in this case of over 50%. For cases in which only lower frequencies are important, the corresponding

Figure 4.31: Left binaural signal in the time domain (exhaustive method and ESR method with 16 sources)
lower number of equivalent sources can be used with an even greater reduction in source number.

4.3.3 Combined Reduction Methods

Since the ESR method (section 3.3) works to reduce the number of input sources while the SVD method (section 3.1) replaces the HRIR convolution with faster operations, the two methods can be used simultaneously to yield an even greater reduction in computation time. Figures 4.32 and 4.33 show the frequency and time domain comparisons, respectively, of the binaural signals (configuration of figure 4.22) found through the combined reduction methods and the exhaustive method. In this case, three singular values and an eight by eight grid of equivalent sources were used. Because the ESR method for an eight by eight grid is only accurate up to 980 Hz, a low-pass filter at 800 Hz was applied. A correlation of 97% was found between the combined reduction method and the exhaustive method. This level of accuracy is sufficient to conclude that the combined reduction method can be effectively used in this research.
4.3.4 Comparison of Computation Times

Now that it has been verified that the ESR, SVD, and combined reduction methods can accurately replace the exhaustive method, it remains to determine if these methods actually reduce the number of computations in this analysis. To handle this problem, the researchers developed an estimate of the computation time required for each method that is based solely upon the number of additions and multiplications used in each method.

The schematic of the exhaustive method is shown below in figure 4.34. Since the radiation model process is scalar multiplication, the number of computations is modeled to be $N$ multiplications. Because convolution is required in the process of filtering the radiated pressure with the HRTFs, the number of computations required will be $128 \times N$ multiplications plus $128 \times N$ additions (where 128 is the number of samples in the HRIRs from MIT). Similarly, the additions required to sum the signal are counted as $N$ summations. In this model, addition and multiplication are estimated to require
approximately the same amount of computation time. Adding up the number of summations and multiplications required by the exhaustive method yields a total number of computations, \( C_E \):

\[
C_E = 258N
\]  

(4.1)

where \( N \) is the number of sources. In the case of the original plate, with a 23 by 23 grid of measurement points, the total number of computations per iteration would be 136,482.

In the case of the ESR method, the schematic is the same as in figure 4.34, and the number of computations is the same as in equation 4.1. The difference is found in the number of sources (\( N \)). To maintain the 0 to 2000 Hz frequency range of accuracy, a 16 by 16 grid of equivalent sources can be used and reduce the number of computations per iteration to 66048. For a model with 64 sources (which is accurate up to 980 Hz), the number of computations will be 16,512. If only 16 sources are used, the number of calculations would be reduced to 4128, but the data is only accurate up to 490 Hz. There is an obvious reduction in computation time achieved using the ESR method, but in order to achieve the greatest computation reduction there is a tradeoff involving the frequency range of accuracy.

The reduction in computation time achieved with the SVD method is less intuitive because the system is more complex (figure 4.35). The computations required by the radiation model, though, are the same as with the ESR method: \( N \) multiplications. Multiplying by the left singular vector, \( U(i) \), requires \( N \) multiplications for each of the \( M \) singular values, which is \( M*N \) multiplications. By the same reasoning, the summing process requires \( M*N \) summations. Since the convolution occurs after the summing process in the SVD method, only \( M*128 \) multiplications and \( M*128 \) additions are required. The scalar multiplication of the singular values is not included in the number of
computations because it can be multiplied by the right singular vectors in pre-processing. The total number of computations, $C_s$, is:

$$C_s = N + 2MN + 257M$$  \hspace{1cm} (4.2)$$

where $M$ is the number of singular values used in the HRIR reconstruction. If three singular values and 529 input sources are used, $C_s$ will be 4474 computations. According to figure 3.6, the error associated with using three singular values will be -20 dB (i.e., minimal) for a frequency range up to 2500 Hz.

If ESR is combined with the SVD method, the number of input sources can be reduced to 64 sources, for example. In that case, the total number of computations (found by equation 4.2) would be 1219. When compared with the 136,482 computations required by the exhaustive method, (even though this model is only an approximation) it is safe to conclude that the ESR, SVD, and combination methods all produce significant reduction in the required number of computations and in turn, processing time. Table 4.4, below, offers a comparison of the computation time associated with each method.
This reduction in computation time is important in order to create a real-time binaural simulation of structural acoustic data (as in section 2.4). In a real-time virtual environment, the user’s head orientation is measured by a head-tracking unit and fed into the computer. Typical update rates for head-tracking units are less than 100 Hz, corresponding to one update every 10 ms. Until the computer receives another head orientation from the head-tracking unit, the left and right ear signals are calculated based upon the current head orientation. However, the audio and visual outputs are also based on other inputs. For instance, if the simulation were of a jet flying overhead, the sound input from the simulated jet would typically be updated at a sample rate of 44100 samples/second. This means that in order to create a structural acoustics simulation using the exhaustive method (for 529 sources), the computer must make an estimated 136,482 calculations per ear in about 23 $\mu$s. With the combined reduction methods, only 883 calculations are required per ear in the same amount of time. For this number of sources, an estimated computation time reduction of 99.4% can be achieved. This reduction would be even greater for a larger number of input sources, making real-time binaural analysis of structural acoustic data much more feasible.

### 4.4 Validation using Stereo-Dipole Technique

Having mathematically validated the accuracy and efficiency of the exhaustive, SVD, and ESR methods, these methods were also verified using auditory tests. Since mathematical error does not necessarily correspond to auditory errors\textsuperscript{34}, thorough
listening tests using a number of impartial subjects are needed in order to unquestionably verify the exhaustive and reduced methods. In this research, preliminary auditory tests were conducted using the stereo dipole method, and more sophisticated tests are a matter of future work.

These preliminary tests were conducted in an anechoic chamber (figure 4.36) using the stereo-dipole technique described in Appendix 1. The pre-calculated binaural speaker signals were outputted from a laptop computer, passed through an Ithaco 2 channel filter and a Rane MA6 amplifier, and played through two Polk Audio speakers. The speakers were mounted in the Vibrations and Acoustics Laboratory’s anechoic chamber about 12” apart and 4’ from the chamber floor. A chair was positioned in the anechoic chamber approximately 5’ 9” from the front of the speakers in order to create a 5° angle between each speaker and the head center of a listener sitting in the chair. This angle between the speakers and the chair is important in the creation of the speaker filters (see Appendix 1). The auralization setup was tested by playing speaker signals from the Institute of Sound and Vibration Research on the computer that were known to produce a three-dimensional sound field through a stereo-dipole arrangement. The two listeners generally agreed that this setup provided excellent externalization and minimized front-back confusion, but was less effective at producing elevation cues.

Figure 4.36: Setup of stereo-dipole auditory tests
After testing the auditory test setup, the speaker signals resulting from the current research were played from the computer and over the speakers. The speaker signals (as in the schematic of 2.10) that were outputted from the computer consisted of binaural signals that had been filtered through the stereo-dipole speaker filters. Four sets of binaural signals were compared, which were calculated through the exhaustive, SVD, ESR, and combined ESR and SVD methods. Binaural signals were compared for different combinations of azimuth angle, elevation angle, virtual distance between the head and the plate, and type of input force (impulse, random input, or pure tone). In addition, binaural signals were calculated using only one node on the plate in order to test the sensation of a distributed source versus a point source.

Preliminary auditory tests of these binaural signals verified that all four methods (exhaustive, SVD, ESR, and combined methods) provided realistic auralization of the sound of a vibrating plate. Because of the distributed nature of the source, it was sometimes difficult to isolate the exact location of the source, but the source direction could be identified within a range of about $30^\circ$. In addition, the direction of the sounds calculated using only one node of the plate could be clearly pinpointed, and when compared with the sounds for the entire plate, the distributed nature of the plate sounds was clearly distinguishable. In general, both the exhaustive and reduced methods provided the listeners with a three-dimensional audio sensation and there was no difference between the auralization of these methods that could be detected by the subjective listeners.

4.5 Conclusion

The accuracy and effectiveness of the exhaustive and reduced methods were validated through the use of measured plate data that was collected by Grosveld. The exhaustive method results were compared with measured data and found to match sufficiently, despite different estimates of the low damping level of the plate. The accuracy of the ESR method was found to be 98% over the frequency range of accuracy (which depends upon the number of equivalent sources that are used). Similarly, the SVD method was found to be 97.7% correlated to the exhaustive method. While maintaining the accuracy of the exhaustive method, the combined reduction methods
produced a significant computational reduction (99.4% reduction when 16 equivalent sources and 3 singular values replace the 529 source exhaustive method). Preliminary auditory tests were also conducted with the stereo-dipole method and verified that the reduced methods produce the same auralization as the exhaustive method. In summary, the reduced methods can more efficiently produce binaural signals, while maintaining the mathematical and auditory accuracy of the exhaustive method.
CHAPTER 5: CONCLUSION

This thesis has discussed the creation of a binaural simulation of structural acoustic data and the reduction methods that were developed to make that simulation more efficient. After introducing the topic and reviewing current literature surrounding this work, the general theory behind binaural simulation was addressed. This discussion of theory was concluded in the third chapter with a description of the theory behind the reduction methods that were investigated, most notably SVD and ESR. Finally, the fourth chapter presented the results found from implementing the exhaustive and reduced methods with measured data. Each of these chapters will be briefly summarized here and conclusions will be drawn regarding the overall effectiveness of the reduction methods. In addition, future work in the area of structural acoustic binaural simulation will also be discussed.

5.1 Chapter Summaries

The introduction to this thesis began with a discussion of general binaural acoustics, defining Head Related Transfer Functions (HRTFs), Interaural Time Difference (ITD), and Interaural Intensity Difference (IID). Recent literature on binaural acoustics was also discussed, especially efforts to model HRTFs. Although the SVD method has been previously investigated\textsuperscript{12, 13}, it has not yet been applied to a vibro-acoustic source and no results have been published regarding its accuracy or effectiveness. Similarly, the ESR method has been developed by Koopman and others, but has not yet been applied to a binaural simulation.

Following the introduction and review of recent literature, the second chapter offered a description of the general theory behind binaural simulation. A computational plate model was discussed that determines the surface accelerations of a vibrating plate. Sound radiation from a monopole, group of monopole sources, and array of monopoles representing a vibrating structure was also explained. In addition, the application of HRTFs and the exhaustive method of binaural simulation were defined. The last section
of this chapter described the elements that combine to produce a real-time binaural simulation.

In the third chapter, three different reduction methods were presented: the SVD method, wavenumber filtering, and the ESR method. While wavenumber filtering was eliminated as a possible reduction method because it causes imaginary sources to appear outside the edges of the panel and cannot be used for complex shaped sources, the SVD and ESR methods were both found to be promising reduction methods. The ESR method reduces the number of sources by filtering out the non-radiating wavenumber components (as with wavenumber filtering), is able to be used for complex sources, and does not create sources outside the vibrating structure. In addition to the pre-processing ESR method, the real-time SVD method replaces the complex convolution process with several more efficient steps involving scalar multiplication. Since the ESR method reduces the number of sources necessary to represent a vibrating plate, and the SVD method reduces the complexity of the convolution process, the most computationally expensive part of binaural simulation (convolution of a large number of sources with the HRTFs) can hypothetically become much more efficient.

Having developed the theory behind the binaural simulation methods, the accuracy and effectiveness of the exhaustive and reduced methods were implemented and validated through the use of measured plate data that was collected by Grosveld. The radiated pressures and binaural signals calculated through the exhaustive method were compared with measured signals and found to match sufficiently, despite different estimates of the low damping level of the plate. The correlation between the ESR method’s binaural signals and those of the exhaustive method was found to be 98% over the frequency range of accuracy (such that the source spacing equals half an acoustic wavelength). Similarly, the binaural signals found through the SVD method were found to be 97.7% correlated to those of the exhaustive method. The combined reduction methods also produced a significant computational reduction (99.4% reduction when 16 equivalent sources and 3 singular values replace the 529 source exhaustive method). Preliminary auditory tests were also conducted with the stereo-dipole method and verified that the reduced methods produce the same auralization as the exhaustive method. In summary, the combined reduction methods can efficiently create binaural
signals that accurately reproduce the binaural signals of the exhaustive method and provide a sensation of auralization to a listener.

5.2 Future Work

In order to further support the accuracy of the SVD and ESR methods, more sophisticated auditory tests could be carried out in the future. Such tests would include several impartial subjects that would each be asked to determine the direction and distributed nature of each sound in a randomized set of binaural recordings of a vibro-acoustic source. The data collected from these subjects would be used to quantify the overall effectiveness of each binaural calculation method. In addition, the effectiveness of each method in handling front-back confusion, higher and lower elevations, and externalization (the sensation that the sound emanates from outside the head, rather from within) could all be addressed. Through this type of auditory test, a measure of the effectiveness of the reduction methods could be found which represents the 360° auditory space.

These listening tests would be even more conclusive if individualized HRTFs were used. The HRTFs for each listener could be measured and each individual’s HRTFs could be used to create an individualized binaural simulation for each listener. Since several errors, such as front-back confusion, are thought to originate from the use of generalized HRTFs, the use of individualized HRTFs would minimize those errors. The errors due to the reduction methods could then be easily distinguished.

In addition to auditory tests, this work could also be supplemented by including sound reflections in the binaural simulation. Theory indicates that both the SVD and ESR methods will remain valid with the addition of a reflective surface, but the ESR method could become more complicated, because the $\mathbf{T}(\omega)$ matrix (equation 3.9) will be more difficult to compute. Through these reflective tests, the use of image sources to account for reverberation could also be tested. In addition, measured data including a reflective surface has also been collected by Grosveld5, and is available for implementation (see Appendix 7).

Another implementation that would complement this work would be different surface geometries. As with the reflective surface, the SVD method will not change for a
complex geometry. In addition, the theory for the ESR method has already been developed for an enclosure and a general shaped structure. However, surface accelerations and transfer functions have not been gathered for the case of a complex shaped structure. Nonetheless, implementation of the reduction methods for a complex geometry and analysis of the results would enhance the findings of this research.

Another addition to this work would be analysis of the reduction methods at higher frequencies. The measured data used in this research is accurate up to 2000 Hz, enough to compare both the ITD and IID of the exhaustive and reduction methods (the primary auralization cue is ITD up to 1000 Hz, after which IID becomes the dominant cue). However, the higher frequencies do contribute to a full auralization experience (their effect is to mainly simulate the pinnae resonance) and this field would benefit from future work comparing the results over a higher frequency range.

This field could also be furthered by a full implementation of both the SVD and ESR reduction methods in a real-time binaural simulation (as in figure 2.10) of structural acoustic data. Such an implementation would provide full proof that the SVD and ESR methods accurately represent the three-dimensional sound of a vibrating plate. In addition, more quantitative comparisons could be made regarding the computational reduction achieved by the SVD, ESR, and combined methods. A work of this type is being undertaken at NASA Langley Research Center by Stephen A. Rizzi and Brenda M. Sullivan in order to represent the three-dimensional sound field produced by the vibrating walls of the space station.50, 51

In addition to further confirming the accuracy and computational effectiveness of the SVD and ESR methods, the SVD method could potentially be used as a tool for understanding the acoustic properties of HRTFs. Nelson and Kahana15 have found that the singular value decomposition of green’s functions from a sphere to a measurement point in space is related to spherical harmonics by a matrix transformation and have also conducted investigations into singular value decomposition of sound radiated from an ellipsoid and simplified pinnae. Since the findings of their research indicate that there is a connection between the mathematical SVD and the acoustically relevant spherical harmonics, the SVD of the HRTFs may provide meaningful insight into the acoustic construction of HRTFs. In order to accomplish this, the SVD would have to be
performed on a very fine, uniform grid of measured HRTFs (unlike the set of HRTFs from MIT6).

While the future work in this area includes auditory tests, implementation using a reflective or complex-shaped surface, analysis over a higher frequency range, completion of a real-time binaural simulation of structural acoustic data, and using SVD to find the acoustical properties of HRTFs, the current work has provided several advancements to this field of research. The exhaustive method of binaural simulation of structural acoustic data has been developed as a benchmark against which to compare the reduction methods. Also, the SVD method has been applied to a vibro-acoustic source, which has not yet been discussed in current literature. Similarly, the ESR method developed by Koopman and others has been applied to a binaural simulation. After using measured structural acoustic data to implement the ESR and SVD methods in binaural simulation, the results have been quantitatively compared with the exhaustive method of binaural simulation. Finally, the number of computations involved in each method have been estimated and compared. Having found that the SVD and ESR methods are computationally efficient, mathematically accurate, and acoustically sound, they were deemed acceptable methods to replace the exhaustive method of binaural simulation.
Literature Cited


10 Chen, J., B.D. Van Veen, and K.E. Hecox, “A spatial feature extraction and
regularization model for the head related transfer function,” *Journal of the

based on spatial feature extraction for the head-related transfer function,” *Journal

12 Abel, J.S., and S.H. Foster, “Method and apparatus for efficient presentation of high-

13 Abel, J.S., and S.H. Foster, “Method and apparatus for efficient presentation of high-
quality three-dimensional audio including ambient effects,” US Patent 5,802,180

14 Evans, M.J., J.A.S. Angus, and A.I. Tew, “Analyzing head-related transfer function
measurements using surface spherical harmonics,” *Journal of the Acoustical

the head-related transfer function,” *Journal of Sound and Vibration*, 239(4)

16 Runkle, P.R., M.A. Blommer, and G.H. Wakefield, “A comparison of head related
transfer function interpolation methods,” *IEEE ASSP Workshop on Applications
of Signal Processing to Audio and Acoustics* (Oct. 1995).

17 Jenison, R.L., “A spherical basis function neural network for pole-zero modeling of
head-related transfer functions,” *IEEE ASSP Workshop on Applications of Signal
Processing to Audio and Acoustics* (Oct. 1995).

using a logarithmic error measure,” *IEEE Transactions on Signal Processing*,

19 Georgiou, P.G., and C. Kyriakakis, “Modeling of head related transfer functions for
immersive audio using a state-space approach,” *Conference Record of the Thirty-
Third Asilomar Conference on Signals, Systems & Computers*, 1 (1999), pp. 720-
724.


40 Nishino, T., S. Kajita, K. Takeda, and F. Itakura, “Interpolating Head Related Transfer Functions in the median plane,” *Proceedings of the 1999 IEEE Workshop on*
Applications of Signal Processing to Audio and Acoustics, New Paltz, New York (October 17-20 1999).


   [www.isvr.soton.ac.uk/FDAG/vap/](http://www.isvr.soton.ac.uk/FDAG/vap/)


InterTrax 2, *InterSense*, [www.isense.com](http://www.isense.com)


Once a binaural signal is calculated, it must somehow be transmitted to the listener’s ears. The ideal transmission method would perhaps be to place speakers inside the ears of a listener. This method would provide a realistic auralization experience, but has other disadvantages, namely listener discomfort, high cost, and inability to produce low frequency sound (due to the small size of the speaker). The next best transmission method would appear to be headphones (which will be discussed in Appendix 3). While the headphone transmission method is advantageous in many ways, it requires the use of a head-tracking unit in order to produce the appropriate signal when the listener’s head rotates. In addition, many listeners sense that the binaural sound seems to come from inside their head when listening to binaural signals over headphones (i.e., poor externalization). Another way of transmitting sound to a listener’s ears is through loudspeakers placed away from the body.

The stereo-dipole is one of many ways of recreating stereophonic sound at the ears of a listener by transmitting sound from a number of speakers placed at a distance from the listener\textsuperscript{52}. The primary differences between these methods are number of speakers and the location of the speakers with respect to the listener, but all of these methods rely upon the principles of cross-talk cancellation. In cross-talk cancellation, the speaker signals are specially filtered to cancel out the natural sound filtering that occurs as the sound travels from the speaker to the listener’s ears.

In the stereo-dipole method, the speakers are arranged side by side such that the angle between the center of the listener’s head and each speaker is $5^\circ$ (as in figure A1.1). This speaker arrangement was developed at the ISVR\textsuperscript{53,54} and was optimized in order to have a large sweet spot and to be robust with respect to head rotation. (The “sweet spot” refers to the small area in which the listener can experience convincing virtual sound.)
Because both of the speakers send sound to each of the left and right ears (figure 1.2), cross-talk occurs (i.e. sound transmitted from the left speaker to the right ear is cross-talk, and vice versa). In general, cross-talk refers to sound that is heard at the left or right ears that is not intended, so the filtering of sound as it travels from the speakers to the ear (the HRTF for 5° azimuth, 0° elevation) is also cross-talk.\footnote{1}

In order to produce the binaural signals at the ears of the listener, the cross-talk must be cancelled out by pre-filtering the signals. These pre-filters are designed to be the
inverse of the transfer functions from the loudspeakers to the head, that is, inverse HRTFs. The matrix of inverse transfer functions, $W$, is calculated according to equation A2.1:

$$W = \left[ H_0^T + \beta I \right]^{-1} H_0^T$$  \hspace{1cm} (A2.1)

where $I$ is the two by two identity matrix, $\beta$ is a regularization parameter, and $H_0$ is the matrix of speaker HRTFs. $H_0$ is defined according to equation A2.2:

$$H_0 = \begin{bmatrix} H_{5R} & H_{5L} \\ H_{5L} & H_{5R} \end{bmatrix}$$  \hspace{1cm} (A2.2)

where $H_{5L}$ and $H_{5R}$ refer to the HRTFs for 5° azimuth and 0° elevation for the left and right ears, respectively. The regularization parameter, $\beta$, is used to prevent the ill-conditioning found in the inverse of $H_0$ at some frequencies. Although the optimal regularization parameter varies with frequency, a value of 0.01 was used for the stereo-dipole tests discussed in section 4.4.

Once the matrix of speaker filters ($W$) are found, they can easily be applied to the left and right binaural signals and outputted through the speakers in order to produce the binaural signals at the ears of the listener. Figure A1.3 shows the process of applying the speaker filters to the binaural signals, in which the speaker filters 1 and 2 are $W(1,1)$ and $W(1,2)$, for diagram simplicity. (Note that this is valid because $W$ is a symmetric matrix.) This diagram is an expanded view of the speaker filter application, which occurs in figure 2.10 in the real-time binaural simulation process.

![Figure A1.3: Speaker filter application process for the stereo-dipole method](image-url)
One disadvantage of the stereo-dipole method is that its effectiveness is somewhat affected by the environment of the stereo-dipole arrangement.\textsuperscript{1,56} If the stereo speakers are arranged in a reverberant room, additional cross-talk (unwanted sound) will approach the listener’s ears. This causes error in the stereo-dipole method, which is dependent upon the reverberation characteristics of the room. If the speakers are arranged within an anechoic chamber, this error is minimized.
Appendix 2: Image Sources

Consider a two-dimensional four-walled enclosure that contains a source of sound (figure A2.1) and an observer. The highest magnitude sound that will be detected by the observer will be the sound directly emanated from the sound source. However, the sound radiated from the source will also reflect off the walls of the enclosure.

Figure A2.1: Enclosure containing a source of sound and an observer

Figure A2.2 shows the first order reflections that will be detected by the observer. First order reflections are sound rays that are reflected off the side of the enclosure only once before arriving at the detection point. The number of first order reflections is equal to the number of walls in the enclosure, four in this example. Additionally, if the walls are considered to be purely rigid, the magnitude of each reflected wave would not be diminished by its interaction with the wall. (The distance traveled by the reflected ray on its path between the source and the observer would cause the only gain reduction, which is found according to equations 2.8 and 2.9.)

Realistically, each surface has a sound absorptivity that also diminishes the magnitude of the reflected sound. With second and third order reflections (in which the sound rays reflect off the enclosure walls two or three times, respectively) the sound magnitude will
be greatly reduced due to the high number of interactions with the enclosure walls. Reflections of very high orders can be ignored because their magnitude is extremely low compared with the direct sound ray and low order reflections.

![Figure A2.2: Four first order reflections from the sound source to the observer](image)

In order to computationally simulate the reflection of sound inside an enclosure, the path of each reflected sound ray can be traced, accounting for the distance of each ray and the sound magnitude decrease due to reflection off absorptive walls. Rather than use these ray tracing methods to detect the sound at the observer due to the direct and reflected waves from each of the monopoles, image sources\textsuperscript{36, 58, 59} can be used. In certain cases, the use of image sources can offer advantages in computational efficiency. Image sources can be combined with the ESR method of section 3.3 to efficiently create the binaural signals to represent the sound inside an enclosure with vibrating walls.

Figure A2.3 shows the image sources for the four reflected sound rays of figure A2.2. The image sources are positioned such that the distance between the image source and the edge of the enclosure is equal to the distance from the source to the reflection wall. In this way, the time delay and attenuation achieved by the reflection will be represented through the image source. Similarly, the position of the image source is such that the angle of the sound ray approaching the measurement point from the image source will be the same as the angle of the original reflected ray. In addition, the magnitude of the image sources will be less than that of the original source in order to simulate the sound decrease due to sound absorption during reflection. Second and third order (and higher) reflections are represented by corresponding order image sources, which are each placed at successively further distances from the source. As with first order reflections,
the magnitude of the image sources are adjusted so that the magnitude and angle of approach of the sound rays from the image sources will match those of the original reflected rays.

![Image Source Diagram](image-source-diagram.png)

**Figure A2.3:** Image sources produce the same sound at point R as the original reflected waves

As long as the pressure at the observer due to the image sources matches the pressure at the observer found with ray tracing methods, the positions of the image sources can be changed. An example of this is shown in figure A2.4, in which the original image sources are replaced with image sources along the edge of the enclosure. The pressure at the observer can be matched with that due to the original image sources by decreasing the magnitude of the source and adding a time delay to the source signal. This time delay and magnitude shift compensate for the shorter distance that is traveled by the sound from the replacement source. However, the positioning of these replacement image sources will change as the observer changes position (but do not change due to head rotation). This is unlike the original image sources, which accurately reproduce the sound field regardless of the position of the listener.
Instead of using replacement image sources, equivalent source reduction can be used to accurately reproduce the entire sound field created due to direct and reflected waves. The procedure is the same as with a vibrating structure. The velocity normal to an evaluation surface and due to the original source and image sources is matched with the velocity profile created by equivalent sources. The equivalent sources will differ from the replacement sources in two ways. First, the replacement sources are created simply by adding a simple time delay or change in magnitude. In contrast, the equivalent sources will have an entirely different frequency spectrum than the image sources. Second, the equivalent sources will accurately reproduce the sound field regardless of the listener position (as long as the listener remains inside the evaluation surface) and replacement image sources will only produce an accurate sound field for one discrete position. Because ESR accurately reproduces the entire sound field inside the evaluation surface and requires fewer sources than the image sources, replacement of image sources
with equivalent sources is a very promising method of producing accurate reverberation in a binaural simulation.
Appendix 3: Headphone Simulation

Similar to the stereo-dipole method discussed in Appendix 1, headphone simulation is an arrangement for transmitting sound from a set of speakers to the ears of listener. Unlike the stereo-dipole method, the binaural signals are played directly over the headphones because the speakers are assumed to be sufficiently close to the ears that the sound is not filtered as it travels from the headphones to the ears. In addition, there is no cross-talk due to speaker placement because the left ear signal travels only to the left ear, and similarly with the right ear.

While headphone simulation has the advantage of requiring no signal processing, it has other disadvantages. The primary disadvantage to headphone simulation is that listeners experience a lack of externalization.¹ That is, the sounds appear to emanate from inside the listener’s head. This is partially caused by each of two factors: the closeness of the speakers to the head, and by the fact that the headphone simulation is not robust when the head is rotated. This head rotation is important because humans often verify the location of a sound by slightly rotating their head.

The headphone simulation can be improved by the use of a head-tracking device that detects the angular orientation (yaw, pitch, and roll) of the listener’s head and adjusts the binaural signals accordingly. An even greater improvement can be achieved through a head tracker that detects the angular orientation and the position of the listener in x, y, and z coordinates, but such a head tracker is quite expensive. In addition, the use of a head-tracking unit complicates the binaural simulation process.

A head-tracking unit can also be combined with the stereo-dipole method to allow the listener to change his or her xyz position while maintaining the appropriate binaural signals in the listener’s ears. For a relatively small listener motion, the speaker filters can be adjusted by applying a simple time delay to one speaker and detracting the same time delay from the other speaker. For greater listener motion, the speaker filters must be recalculated in real-time using the appropriate HRTF for the listener’s new position, which is a computationally expensive process.
Overall, headphone simulation is a simpler method and can easily be coupled with an angular position head tracker to maintain robustness as the listener’s head rotates. However, the stereo-dipole method provides a more realistic virtual environment because the problem of externalization is solved. The method of choice depends upon the application.
Appendix 4: Real-time headphone simulation

Björn Knöfel and Marty Johnson have created a real-time binaural simulation using the headphone transmission method that can create and display up to four virtual sources at once (note that these sources are point sources, rather than vibro-acoustic, distributed ones). This simulation was implemented using C code and the InterTrax 2 head-tracking unit developed by InterSense.

The schematic of the real-time headphone simulation is shown in figure A4.1. For this simulation, the binaural calculations were found through the exhaustive method of sections 2.2 and 2.3. Real-time simulation requires the binaural signals for one iteration to be calculated while the listener hears the binaural signals that were calculated for the previous iteration. This causes a slight error in the simulation because the binaural signals for the previous iteration are based upon the previous iteration’s head orientation. This causes problems in the simulation if the head is rotating very quickly, but most often this discrepancy is not noticeable because each iteration is only about 12 ms long.

Another problem found in the real-time binaural simulation was a “clicking” sound that was transmitted through the headphones. This phenomenon was caused by sudden changes in the binaural signals due to changes in the listener’s angular position. Fading between the signals from the previous and current iterations solved this problem,
where fading is defined as decreasing the magnitude of the previous signal while increasing the magnitude of the current signal.

The resulting binaural simulation was very realistic. Sounds produced through this simulation were more externalized than those produced from a simulation that did not include a head-tracking unit. As expected, when the head rotated very quickly, there were slight discrepancies that were noted, but otherwise the real-time headphone simulation accurately re-created a three-dimensional sound field.
Appendix 5: Time Delay Removal

As mentioned in the discussion of the SVD reduction method in section 3.1.1, the SVD of the HRTFs can be performed with greater accuracy if the interaural time delay (ITD) for each azimuth and elevation angle is first removed from the HRTFs. This appendix will describe two different methods of removing the delay from the HRTFs.

ITD is a short time delay (a few samples) at the beginning of the HRIRs that varies in length with elevation and azimuth angles. Although the delay is easily visible in a 3D plot of HRIRs (figure A5.1), the delay is not defined by a mathematical equation. The delay can be estimated, though, by determining the first sample number that is equal to or greater than $5\%^{16}$ or $10\%^{10}$ of the value of the maximum magnitude of that HRTF.

![Figure A5.1: HRIRs for 0° elevation and -180° to 180° azimuth](image)

Figure A5.1: HRIRs for 0° elevation and -180° to 180° azimuth
The time delay for 0° elevation and -180° to 180° azimuth that was found using the 5% rule is shown in figure A5.2. Notice that the time delay found using this method is not a very smooth function of angle because the time delay is defined in terms of whole samples, and the sample resolution is too large to smoothly represent the ITD function. Because the ITD function is not smooth, the HRIRs without ITD will not be a smooth function of angle. The SVD of the HRIRs is more accurately represented by fewer singular values (and singular vectors) when the HRIRs are a smooth function of angle.

![Figure A5.2: ITD found using the 5% rule (for 0° elevation HRIRs)](image.png)

This problem can be remedied by representing the ITD as a continuous function of angle, and thereby representing the ITD in terms of partial samples. A sixth order polynomial fit was applied to the ITD found through the 5% rule (shown in figure A5.3). This continuous delay cannot be removed from the HRIRs in the time domain without oversampling the HRIRs, removing the delay, and resampling the HRIRs at the original sample rate (44.1 kHz). This can be more easily accomplished by removing the ITD from the HRTFs in the frequency domain through the following equation:
\[ HRTF_{nd}(\theta, \phi) = HRTF(\theta, \phi)e^{\frac{jt_d}{sr}} \]  

(A5.1)

where \( HRTF_{nd}(\theta, \phi) \) is the HRTF (in the frequency domain) for azimuth angle, \( \theta \), and elevation angle, \( \phi \), with no interaural time delay; \( HRTF(\theta, \phi) \) is the original HRTF for angles \( \theta \) and \( \phi \); \( t_d \) is the interaural time delay (in samples), and \( sr \) is the sample rate (44100 Hz in this case). The resulting continuous function of the ITD is shown in figure A5.3.

![Figure A5.3: Polynomial fit of the ITD found using the 5% rule (for 0° elevation HRIRs)](image)

The ITD that is removed from the HRIRs before the SVD is performed on the HRIRs must still be included in the calculation of the binaural signals in order to accurately produce the three-dimensional sound field. This delay is simply added to the \( r/c \) delay due to sound radiation over the distance \( r \) (equations 2.8 and 2.9). The continuous function of delay can easily be added to the \( r/c \) delay in the frequency domain, but problems arise in the time domain. In order to add in the partial sample delays in the time domain, the signals must be oversampled, shifted in time, and resampled at the original sample rate. This process is very computationally expensive.
In fact, the continuous delay cannot be efficiently used in the time domain because the extra computations necessary to apply partial sample delays outweigh the potential increase in efficiency offered by using the continuous delay in the SVD method.

Although the binaural signals can be calculated in either the time or frequency domains, the application for this work is real-time binaural simulation, which is more efficiently performed when the entire process is in the time domain (as mentioned in section 2.4). Since use of the continuous delay function is not efficient in the time domain, whole sample delays were used (that is, the original ITD function found through the 5% rule and shown in figure A5.2).
Appendix 6: Wavenumbers

There are several types of wavenumbers, including acoustic, bending, and modal wavenumbers, which will all be introduced here. Following these introductions will be a discussion of wavenumber decomposition, specifically why waves whose wavenumber is greater than the acoustic wavenumber will not radiate sound into the far field.

Every type of sound wave has a wavenumber, $k$, which is defined to be:

$$k = \frac{2\pi \lambda}{c_{ph}} \tag{A6.1}$$

where the wavelength, $\lambda$, is the spatial period illustrated in figure A6.1\textsuperscript{32}, and $c_{ph}$ is the phase velocity of the wave. In a fluid, the phase velocity is simply the speed of sound in that fluid medium, so the acoustic wavenumber, $k_a$, is defined as:

$$k_a = \frac{\omega}{c} \tag{A6.2}$$

where $c$ is the speed of sound. The bending wavenumber in a plate is known to be:

$$k_b = \left( \frac{m \omega^2}{D} \right)^{1/4} \tag{A6.3}$$

where $m$ is the mass of the plate, and $D$ is the bending stiffness of the plate, defined as:

$$D = \frac{E h^3}{12(1 - \nu^2)} \tag{A6.4}$$

where $E$ is the modulus of elasticity, $h$ is the plate thickness, and $\nu$ is the Poisson’s ratio of the plate.

![Figure A6.1: Definition of wavelength\textsuperscript{32}](image-url)
To determine whether a bending wave in the plate will radiate sound into the far field, one must know the component of the acoustic wavelength in the $y$ direction. The relationship between the $x$, $y$, and $z$ components of the acoustic wavenumber is:

$$k_x^2 = k_y^2 + k_z^2$$  \hspace{1cm} (A6.5)

where $k_x$, $k_y$, and $k_z$ are the respective components of the acoustic wavenumber, as shown in figure A6.2.

![Diagram of acoustic wavenumber components](image)

Figure A6.2: Relationship between $x$, $y$ and $z$ components of the acoustic wavenumber

In the case of a vibrating plate, $k_x$ defines sound that radiates along the plane of the plate and $k_y$ defines sound that radiates perpendicular to the plane of the plate. Since the bending waves on the plate are creating the acoustic waves,

$$k_x = k_b$$  \hspace{1cm} (A6.6)

In order to find the $y$ component of the acoustic wavenumber, equation A6.6 is substituted into equation A6.5 to yield:

$$k_y = \pm \sqrt{k_a^2 - k_b^2}$$  \hspace{1cm} (A6.7)

The resulting value for $k_y$ will depend upon the relative magnitudes of $k_a$ and $k_b$. If $k_b < k_a$, then $k_y$ will be real, and the angle of the plane sound waves, $\phi$, radiated from the plate will be given by:

$$\cos \phi = \frac{k_y}{k_a} = \sqrt{1 - \left(\frac{k_b}{k_a}\right)^2}$$  \hspace{1cm} (A6.8)
Since the plate cannot generate waves towards itself, the negative sign in equation A6.7 is disallowed.

If \( k_b > k_a \), then \( k_y \) will be imaginary, meaning that evanescent waves are created in fluid in the near-field region. These waves exponentially decay in magnitude and do not radiate sound to the far-field. If \( k_b = k_a \), then \( k_y = 0 \), but the boundary conditions cannot be satisfied because finite vibration yields infinite pressure, which cannot occur physically.\(^3\)

In summary, only bending wavenumbers less than the acoustic wavenumber will radiate sound to the far-field region because the \( y \) component of the acoustic wavenumber will be imaginary if the bending wavenumber is greater than the acoustic wavenumber.
Appendix 7: Additional Measured Data

There are two types of measured data that were collected by Grosveld\textsuperscript{5} and not discussed in chapter 4. In addition to the structural, acoustic, and binaural transfer functions collected to describe the vibrating panel secured in an infinite baffle, acoustic and binaural transfer functions were also collected with a reflective panel secured to the baffled panel. A B-format microphone was also used to record the three-dimensional sound field in the tests both with and without the reflective surface. Although these two types of measured data were not directly used in this research, they are described here because they may be used in future work.

Figure A7.1 shows the KEMAR manikin mounted in the anechoic chamber in front of the baffled panel (on the right) and the reflective panel (on the left). The aluminum reflective panel measured 1473 by 1473 mm and was installed along the left vertical edge of the baffled panel such that the two panels were perpendicular to each other. Structural transfer functions were not collected for the reflective panel, but acoustic and binaural transfer functions were acquired. These transfer functions were measured at locations 1 and 2 of table 4.2. Measurements were not acquired at location 3 because the position was too far from the reflective surface to detect any reflected sound. In future work the data measured with this reflected panel could be used to verify the exhaustive and reduced methods ability to simulate a reflective room.\textsuperscript{33}
A B-format microphone was also used to measure the sound field due to the baffled and reflected panels. Figure A7.2 shows the KEMAR manikin and a Soundfield\textsuperscript{61,62} B-format microphone positioned in front of the two panels. The B-format microphone actually consists of 4 microphones, which are a reference microphone and directional microphones in the x, y, and z directions. Together these four signals represent the entire three-dimensional sound field. Specifically, these microphones capture the zero and first order spherical harmonics in three directions.\textsuperscript{5} The B-format microphone was oriented with its manufacturer logo facing the baffled panel.\textsuperscript{62}
The B-format recordings were performed for the purpose of verifying the binaural calculations. The B-format representation of a reference and three vectors is mathematically simple to reproduce and can easily be translated into other formats. Since the B-format signals represent the entire three-dimensional field at that point in space, these signals can be combined with HRTFs to represent the binaural signals for any head rotation about that point in space. In future work, the B-format signals could be used to verify the measured and calculated binaural signals at the ears of the KEMAR head.
Vita

Aimee L. Lalime was born in Vancouver, WA on August 9, 1977 to Joanne and Manuel Costa. She graduated *Cum Laude* with a Bachelor’s of Science degree in Mechanical Engineering from Virginia Polytechnic Institute and State University in December, 2000. Immediately following the completion of her Bachelor’s degree, she began to pursue her Master’s of Science degree in Mechanical Engineering with a concentration in acoustics, also at Virginia Tech. After the completion of her Master’s degree in August, 2002, she will begin working as a Mechanical Engineer in Los Angeles, CA.