A NEW METRIC TO PREDICT LISTENER ENVELOPMENT BASED ON
SPHERICAL MICROPHONE ARRAY MEASUREMENTS AND
HIGHER ORDER AMBISONIC REPRODUCTIONS

A Dissertation in
Acoustics
by
David A. Dick

© 2017 David A. Dick

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

December 2017
The dissertation of David A. Dick was reviewed and approved* by the following:

Michelle C. Vigeant  
Assistant Professor of Acoustics and Architectural Engineering  
Dissertation Advisor  
Chair of Committee

John F. Doherty  
Professor of Electrical Engineering

Daniel A. Russell  
Professor of Acoustics  
Director of Distance Education for the Graduate Program in Acoustics

Victor W. Sparrow  
Professor of Acoustics  
Director of the Graduate Program in Acoustics

Bill Rabinowitz  
Manager of Acoustics Research, Bose Corporation  
Special Member

*Signatures are on file in the Graduate School
ABSTRACT

The objective of this work was to create a new metric to predict listener envelopment (LEV), the sense of being surrounded by the sound field, based on 32-channel spherical microphone array measurements taken in a number of venues and a series of listening tests. A spherical microphone array was used to investigate LEV because it can be used for a) high resolution spatial analysis of the sound field in full 3D via beamforming techniques, and b) subjective listening tests using 3D reproductions of the sound fields over a loudspeaker array via Ambisonics. This work is comprised of three separate studies: a first study validating the spherical microphone array measurement system, a second study investigating LEV in a 2,000-seat concert hall, and a third study in which a new metric is proposed to predict LEV based on listening tests using measurements obtained in seven additional halls.

A study was conducted to validate the spherical microphone array measurement system. Spatial room impulse response (IR) measurements were taken in a 2500-seat auditorium to determine how room acoustic metrics measured with a spherical microphone array compare to those measured with the traditional microphone setup (an omnidirectional and figure-8 microphone pair). Measurements were obtained at six receiver locations with three repetitions each to evaluate repeatability. The metrics considered in this study were: reverberation time (T30), early decay time (EDT), clarity index (C80), strength (G), lateral energy fraction (JLF) and late lateral energy level (L). For the spherical array measurements, the omnidirectional (monopole) and figure-8 (dipole) patterns were extracted via spherical harmonic beamforming. The measurements were found to be consistent both across repetition and microphone configuration. The results from this study indicate that spherical microphone arrays can be used to both measure existing LEV metrics, and to develop a new metric to predict LEV.

An LEV study was conducted using spherical microphone array IRs obtained in a 2,000-seat concert hall in several receiver locations and hall absorption settings. The IRs were analyzed using a 3rd order plane wave decomposition (PWD) beamformer. Additionally, the IRs were convolved with anechoic music and processed for 3rd order Ambisonic reproductions and
presented to subjects over a 30-loudspeaker array. Instances were found in which the energy in the late sound field did not correlate with LEV ratings as well as energy in a 70 to 100 ms time window. Follow-up listening tests were conducted with hybrid IRs containing portions of a highly enveloping IR and a highly unenveloping IR with crossover times ranging from 40-140 ms. Additional hybrid IRs were studied wherein portions of the spatial IRs were collapsed into all frontal energy with crossover times ranging from 40-120 ms. The tests confirmed that much of the important LEV information exists in the early portion of these IRs.

In a final LEV study, spherical microphone array IRs were obtained in seven additional halls of various sizes and shapes. The IRs were used for listening tests that included stimuli that were presented as-measured, which included level differences, stimuli that were equalized for level differences, and hybrid stimuli generated by combining portions of enveloping IRs and unenveloping IRs. A new metric was developed named mid-late spatial energy, $J_S$, by integrating energy from a 3rd order PWD of the room IRs as a function of frequency, azimuthal angle, elevation angle, and time, and adjusting the integration limits to maximize correlation between integrated energy and LEV ratings. The difference in overall level between halls was found to be highly correlated with the perception of LEV, but for level-equalized stimuli the correlation was maximized by integrating energy from 60 ms to 400 ms, rejecting sound from the front ±20° in azimuth and rejecting sound from ±70° in azimuth behind the listener. This new metric has a higher correlation with LEV ratings than the currently used metric of late lateral energy level, $L_J$. 
TABLE OF CONTENTS

List of Figures ........................................................................................................................................... xi

List of Tables .................................................................................................................................................. xix

Acknowledgements......................................................................................................................................... xx

Chapter 1: Introduction .................................................................................................................................. 1

1.1 Listener Envelopment (LEV) ................................................................................................................... 1
  1.1.1 Early History of LEV .......................................................................................................................... 1
  1.1.2 Impact on LEV from energy other than late lateral arrivals ............................................................ 4
  1.1.3 Other Proposed LEV Metrics ............................................................................................................ 5
  1.1.4 LEV Summary .................................................................................................................................... 7

1.2 Spherical Microphone Arrays .................................................................................................................. 8
  1.2.1 Wave equation in spherical coordinates ............................................................................................. 8
  1.2.2 Spherical Fourier Transform of Microphone Array Signals ............................................................ 12
  1.2.3 Scattering off of a rigid sphere ......................................................................................................... 13
  1.2.4 Beamforming ..................................................................................................................................... 15

1.3 Ambisonics .............................................................................................................................................. 17
  1.3.1 Spherical Harmonics Format ............................................................................................................ 18
  1.3.2 Ambisonic Decoding .......................................................................................................................... 18
  1.3.3 Max-RE Decoding ............................................................................................................................. 19
  1.3.4 Nearfield Compensation .................................................................................................................... 21
  1.3.5 AURAS Facility .................................................................................................................................. 22
    1.3.5.1 Layout ........................................................................................................................................ 22
    1.3.5.2 AURAS Ambisonic Decoder ...................................................................................................... 22
    1.3.5.3 Ambisonic Performance ............................................................................................................. 23

1.4 Dissertation Outline .................................................................................................................................. 25

Chapter 2: A Comparison of Measured Room Acoustics Metrics Using a Spherical Microphone Array and Conventional Methods ........................................................................................................... 27
2.1 Introduction ...........................................................................................................................................28
2.2 Room Acoustics Metrics .....................................................................................................................30
2.3 Measurement Uncertainty ..................................................................................................................31
2.4 Spherical Microphone Array Processing and beamforming ................................................................33
2.5 Measurements ......................................................................................................................................36
  2.5.1 Measurement Hardware ..............................................................................................................36
  2.5.2 Anechoic Chamber Directivity and Frequency Response Measurements .....................................38
  2.5.3 Impulse Response Measurements in Eisenhower Auditorium .....................................................40
2.6 Data Processing ....................................................................................................................................42
  2.6.1 Microphone frequency response compensation ...........................................................................42
  2.6.2 Rotating the dipole .......................................................................................................................43
  2.6.3 Room acoustic metric calculation ................................................................................................44
2.7 Results ..................................................................................................................................................44
  2.7.1 Measurement Repeatability of the Two Microphone Configurations .............................................44
  2.7.2 Differences in Measured Room Acoustics Metrics between Microphone Configurations .......48
2.8 Conclusions ..........................................................................................................................................53
2.9 Acknowledgements .............................................................................................................................55

Chapter 3: An Investigation of Listener Envelopment Utilizing a Spherical Microphone Array and Third-Order Ambisonics Reproduction ..................................................................................56
3.1 Introduction ..........................................................................................................................................57
  3.1.1 Spherical Array Beamforming ....................................................................................................61
  3.1.2 Ambisonics Reproduction ..........................................................................................................63
3.2 Room Impulse Response Measurements ..........................................................................................64
  3.2.1 Measurement Hardware ..............................................................................................................64
  3.2.2 Room IR Measurements ..............................................................................................................66
3.3 Stimulus Reproduction .......................................................................................................................67
  3.3.1 Auralization and Reproduction of Acoustic Sound-fields (AURAS) Loudspeaker Array ..........67
Chapter 4: A New Metric to Predict Listener Envelopment based on Spatial Impulse Response Measurements and Subjective Listening Tests ................................................................. 90

4.1 Introduction ........................................................................................................... 91
  4.1.1 Spherical Microphone Array Beamforming .................................................... 93
  4.1.2 Ambisonics Reproduction ............................................................................. 94

4.2 Spatial Room IR Measurements ........................................................................... 95
  4.2.1 Measurement Hardware ................................................................................ 95
    4.2.1.1 Spherical microphone array ...................................................................... 95
    4.2.1.2 Three-way omnidirectional sound source ................................................. 95
  4.2.2 Details of the Seven Measured Halls and the Receiver Locations .................. 96

4.3 Subjective Study using Ambisonic Reproductions of spatial IRs ......................... 98
  4.3.1 Subjective listening test stimuli ....................................................................... 99
    4.3.1.1 Test set 1: IRs from Seven different halls as measured .............................. 100
4.3.1.2 Test set 2: IRs from Seven different halls equalized for level ......................................................... 100
4.3.1.3 Test set 3: High-LEV (HLEV) and Low-LEV (LLEV) hybrid IRs .......................................................... 101
4.3.1.4 Test set 4: LLEV&HLEV hybrid IRs .................................................................................................. 102

4.4 Subjective Study Results ......................................................................................................................... 102
4.4.1 Statistical Analysis of LEV Ratings ...................................................................................................... 102
4.4.2 Results: Set 1 – IRs from 7 different halls as measured ........................................................................ 103
4.4.3 Results: Set 2 – IRs from 7 different halls equalized for level .............................................................. 105
4.4.4 Results: Set 3 – H-LEV&L-LEV Hybrid IRs ......................................................................................... 107
4.4.5 Results: Set 4 – L-LEV&H-LEV Hybrid IRs ......................................................................................... 109

4.5 Metric Development ............................................................................................................................... 112
4.5.1 Azimuthal Angular Dependence .......................................................................................................... 113
4.5.1.1 LEV Rating Correlation Excluding Frontal energy ........................................................................... 114
4.5.1.2 LEV Rating Correlation Excluding Rear energy ............................................................................ 114
4.5.2 Elevation Angular Dependence ........................................................................................................... 116
4.5.3 Time Dependence .................................................................................................................................. 117
4.5.3.1 LEV Rating Correlation Varying Early Time Cutoff ....................................................................... 117
4.5.3.2 LEV Rating Correlation Varying Late Time Cutoff ......................................................................... 118
4.5.4 Overall Metric ........................................................................................................................................ 119
4.5.4.1 Performance relative to late lateral energy level and strength ......................................................... 120
4.5.4.2 Performance with order reduction .................................................................................................... 122

4.6 Effect of the Early Sound Field ............................................................................................................... 122

4.7 Conclusions ............................................................................................................................................ 123

Chapter 5: Conclusions ................................................................................................................................. 125
5.1 Results Summary ...................................................................................................................................... 125
5.2 Future Work ............................................................................................................................................. 128

Appendix A: Calculation of Mid-Late Spatial Energy (Js) ............................................................................ 130
Appendix B: Three Way Omnidirectional Sound Source ............................................................................. 132
B.1 Introduction .................................................................................................................. 132

B.2 Directivity of the source components ........................................................................ 136
   B.2.1 Subwoofer Directivity .......................................................................................... 136
   B.2.2 Mid-frequency dodecahedron directivity ............................................................. 136
   B.2.3 High-frequency dodecahedron directivity ............................................................ 138

B.3 Omnidirectional source validation with room measurements .................................... 139
   B.3.1 Measurement uncertainty due to source rotation .................................................. 140
      B.3.1.1 Energy as a function of source rotation ......................................................... 140
      B.3.1.2 Room acoustic metrics as a function of source rotation ............................... 142
   B.3.2 Stacked vs. coincident configurations ................................................................. 144
      B.3.2.1 Energy differences between stacked and coincident configurations .......... 146
      B.3.2.2 Room acoustic metrics for the stacked and coincident orientations ............ 147

B.4 Summary .......................................................................................................................... 149

Appendix C: Complete Set of Directivity Plots of the Mid-Frequency and High-Frequency
Dodecahedrons 151

C.1 High-Frequency Dodecahedron: 3D Octave Band Plots (1,000 – 16,000 Hz) .............. 151
C.2 High-Frequency Dodecahedron 3D Third-Octave Band Plots (1,000 – 16,000 Hz) .... 154
C.3 High-Frequency Dodecahedron: 2D Polar Plots (Azimuthal Plane) Third-Octave Bands (1,000 – 20,000 Hz) .................................................................................................................. 161
C.4 High-Frequency Dodecahedron: 2D Polar Plots (0° Elevation Plane) Third-Octave Bands (1,000 – 20,000 Hz) .................................................................................................................. 164
C.5 High-Frequency Dodecahedron: 2D Polar Plots (90° Elevation Plane) Third-Octave Bands (1,000 – 20,000 Hz) .................................................................................................................. 167
C.6 Mid-Frequency Dodecahedron: 2D Polar Plots (Azimuthal Plane) Third-Octave Bands (125 – 20,000 Hz) .................................................................................................................. 170
C.7 Mid-Frequency Dodecahedron: 2D Polar Plots (Elevation Plane) Third-Octave Bands (125 – 20,000 Hz) .................................................................................................................. 174
LIST OF FIGURES

Figure 1-1: Angular distribution of loudspeakers used to develop $L_J$ (Fig. 1, Bradley and Soulodre, 1995 [10]). .................................................................3

Figure 1-2: Plots of the real-valued spherical harmonics for orders $n = 0, 1, 2,$ and 3.............11

Figure 1-3: Plane wave modal coefficients for a sphere of radius $a = 4.2$ cm .......................15

Figure 1-4: Beam pattern of a truncated plane wave of order $N = 3$. .................................16

Figure 1-5: Block diagram for the overall beamforming system (adapted from Fig. 5.3, Fundamentals of Spherical Array Processing, Rafaely, 2015 [26]). .........................17

Figure 1-6: Energy received at the listening location from a plane wave produced using 3rd order Ambisonics for basic decoding (left) and Max-$r_E$ decoding (right).............21

Figure 1-7: The AURAS loudspeaker array (a), and the distribution of the 30 loudspeakers in the array (b). ................................................................................22

Figure 1-8: Magnitude, direction, and angular error of the $r_v$ vector, plotted in the Ambisonics Decoder Toolbox.................................................................24

Figure 1-9: Magnitude, direction, and angular error of the $r_E$ vector, plotted in the Ambisonics Decoder Toolbox.................................................................25

Figure 2-1: Microphones used in this study. (a) Eigenmike em32 spherical microphone array and (b) a Brüel & Kjær (B&K) Type 4192 omnidirectional and Sennheiser MKH30 figure-8 microphone pair. .................................................................37

Figure 2-2: Custom microphone stand used for accurate and precise placement of microphones. The photo on the left shows the stand being used to place the microphone and the one on the right shows the microphone in the final position. ......38

Figure 2-3: Microphone directivity plots for the Sennheiser MKH30 figure-8 microphone (left) and Eigenmike em32 beamformed dipole pattern (right) ..........................39

Figure 2-4: Deviations from ideal polar patterns: Sennheiser MKH30 figure-8 microphone directivity versus a perfect dipole (a), Eigenmike em 32 beamformed dipole versus a perfect dipole (b) and Eigenmike em32 beamformed omnidirectional pattern versus a perfect omnidirectional (c) pattern .......................40

Figure 2-5: Receiver positions in 2500-seat Eisenhower Auditorium.................................41
Figure 2-6: Eigenmike (blue) and Sennheiser MKH30 (red) equalization filter magnitude response. Target responses are shown as dashed lines, and realized filters fit to the target responses are shown as solid lines. The Eigenmike target and actual filter are nearly identical, which is why the dashed line is difficult to see in the figure. ..................43

Figure 2-7: Standard deviation of the three repeated measurements for each metric at each receiver location for the omnidirectional and figure-8 microphone pair (solid lines) and the Eigenmike array (thin-dashed lines) configurations. The thick-dashed red lines on each plot represent the respective 1 JND for each metric. All standard deviations were found to be well below 1 JND for each metric and all receiver positions, with the exception of a few cases for EDT. ........................................46

Figure 2-8: Differences between the omnidirectional and figure-8 microphone pair and the Eigenmike array configurations for all six metrics measured at all six receiver locations. The thick-dashed red lines on each plot represent the respective 1 JND for each metric. All differences were found to be within 1 JND, with the exceptions of a single point in C80, and several points in EDT at high frequencies. ....................50

Figure 2-9: The energy decay curves and associated early decay time slope fits for R3 in the 2 kHz octave band shown for the omnidirectional B&K microphone (blue) and Eigenmike array (red). Note the spikes in the curve as denoted with the orange ovals, which are likely due to differences in where the centers of each microphone were positioned. ..............................................................52

Figure 2-10: Early decay time calculated for Eigenmike individual microphone capsules (green), and early decay time calculated for the omnidirectional B&K microphone (blue) at R3.................................................................53

Figure 3-1: Spherical harmonic functions of order n and degree m, up to n = 3. For convenience, the real-valued spherical harmonics are shown, where red indicates a positive value and blue indicates a negative value..................................................61

Figure 3-2: Measurement hardware used for the IR measurements: (a) Eigenmike spherical microphone array (mh acoustics em32), (b) B&K binaural mannequin (type 4100-D), and (c) B&K dodecahedron loudspeaker (type 4292)........................................65

Figure 3-3: Receiver locations in the Peter Kiewit Concert Hall. .........................................................66

Figure 3-4: The AURAS loudspeaker array (a), and the distribution of the 30 loudspeakers in the array (b). .....................................................................................................................67

Figure 3-5: Block diagram for Ambisonic reproduction.................................................................68
Figure 3-6: Radial filters convolved with microphone equalization that are applied after encoding the spherical array’s individual microphone signals to Ambisonic signals. ...

Figure 3-7: Comparison of 3rd order simulated plane wave with max-rE decoding, representative of a plane wave produced in the AURAS facility above 1.3 kHz (left) to a plane wave produced in the AURAS facility measured at 2 kHz (right). Pressure magnitude is shown on a linear scale normalized to a maximum value of 1. 

Figure 3-8: Mean LEV ratings for the four test sets. Error bars depict standard errors.

Figure 3-9: Comparison of the late sound field at 1 kHz (80 ms to ∞) between receiver positions with similar LEV ratings (R3 and R8 from Set 2), with 0° azimuth pointing toward the stage and 0° elevation pointing straight up. Sound pressure level is shown ranging from -10 dB to 0 dB, where 0 dB is the maximum level of both sound fields (overall level differences are maintained between the top and bottom plots). The images on the left show the energy distributions over spheres, while the images on the right show the same information, but flattened onto a 2-D plot (similar to an unraveled a map.) Energy at R3 is concentrated toward the front (R3 is underneath a balcony), whereas energy at R8 is more evenly distributed throughout the sphere (R8 is in the top balcony).

Figure 3-10: Comparison of the late sound field at 1 kHz between receiver positions with different LEV ratings (R9 and R10 from Set 1), with 0° azimuth pointing toward the stage and 0° elevation pointing straight up. Sound pressure level is shown ranging from -10 dB to 0 dB, where 0 dB is the maximum level of both sound fields (overall level differences are maintained between the top and bottom plots). The spatial distribution of late energy is similar between the two receivers yet the LEV ratings of these stimuli were significantly different.

Figure 3-11: Comparison of the sound field from 70 to 100 ms at 1 kHz between receiver positions with similar LEV ratings (R3 and R8 from Set 2), with 0° azimuth pointing toward the stage and 0° elevation pointing straight up. Sound pressure level is shown ranging from -10 dB to 0 dB, where 0 dB is the maximum level of both sound fields (overall level differences are maintained between the top and bottom plots). The energy at both receiver positions have a similar level and distribution in terms of lateral, behind, and overhead sound. The frontal energy does differ by 3 dB between the pair, but it is assumed that energy arriving from the front does not influence LEV.

Figure 3-12: Comparison of the sound field from 70 to 100 ms at 1 kHz between receiver positions with different LEV ratings (R9 and R10 from Set 1), with 0° azimuth pointing toward the stage and 0° elevation pointing straight up. Sound pressure level is shown ranging from -10 dB to 0 dB, where 0 dB is the maximum level of
both sound fields (overall level differences are maintained between the top and bottom plots). R9 has a much stronger energy level from behind, which may be contributing to the perceived LEV.

Figure 3-13: Half Hann 2.5 millisecond time windows used to mix two IRs (left), and example resulting hybrid IR (right). The early window and corresponding IR are shown in solid blue, and the late window and corresponding IR are shown in dotted and solid green, respectively.

Figure 3-14: Modified IRs using the early part of R10 (highly unenveloping) and the late part of R3 (highly enveloping) with crossover times ranging from 40 ms (highly enveloping) to 140 ms (highly unenveloping).

Figure 3-15: Modified IRs using the early part of R3 (highly enveloping) and the late part of R10 (highly unenveloping) with crossover times ranging from 40 ms (highly enveloping) to 140 ms (highly unenveloping).

Figure 3-16: Results of the modified IR test in which the early part of the IR is presented in full 3D, and the late part of the IR is reproduced from a single loudspeaker.

Figure 3-17: Results of the modified IR test in which the late part of the IR is presented in full 3D, and the early part of the IR is reproduced from a single loudspeaker.

Figure 4-1: The AURAS loudspeaker array (a), and the distribution of the 30 loudspeakers in the array (b).

Figure 4-2: Three-way omnidirectional sound source components: subwoofer (left), mid-frequency dodecahedron (middle), and high-frequency dodecahedron (right). Note that the photos are not to scale relative to each other. The subwoofer has 25-cm drivers, the mid-frequency source has 10-cm drivers, and the high-frequency source has 1.9-cm drivers.

Figure 4-3: Top-down view diagram of receiver positions measured in each hall. R4, shown as the blue diamond, is the receiver position used for the subjective tests in this study.

Figure 4-4: High-LEV (HLEV) and Low-LEV (LLEV) hybrid IR with a crossover time of 200 ms.

Figure 4-5: LEV ratings for Set 1: IRs from 7 different halls as measured. Colored-shapes were added to indicate statistically significant pairs at $p < 0.05$, where stimuli that share the same colored-shape are not significantly different (note that some data points have multiple colored-shapes).
Figure 4-6: Set 1 LEV ratings vs. late lateral energy level. LEV ratings were found to have a high correlation with $L_J$, although this correlation is primarily due to level differences between the stimuli.

Figure 4-7: LEV ratings for Set 2: IRs from 7 different halls, which are the same halls as used in Set 1, but equalized for level. Colored-shapes were added to indicate statistically significant pairs at $p < 0.05$, where stimuli that share the same colored-shape are not significantly different (note that some data points have multiple colored-shapes).

Figure 4-8: Set 2 LEV rating vs. $L_J$. The correlation between $L_J$ and LEV rating for Set 2 ($R^2 = 0.77$, $p < 0.004$) is lower than in Set 1 ($R^2 = 0.94$, $p < 0.001$) after normalizing the overall A-weighted level.

Figure 4-9: LEV ratings for Set 3. Colored-shapes were added to indicate statistically significant pairs at $p < 0.05$, where stimuli that share the same colored-shape are not significantly different. LEV ratings for crossover times 80 ms and lower are not significantly different than the LEV rating for the whole L-LEV IR.

Figure 4-10: Set 3 LEV rating vs. $L_J$. Each data point is labeled for its crossover time. The correlation between LEV and $L_J$ is high ($R^2 = 0.95$), primarily because all data points up to the 80 ms crossover time share the same $L_J$ value and have similar LEV ratings.

Figure 4-11: LEV ratings for Set 4. Colored-shapes were added to indicate statistically significant pairs at $p < 0.05$, where stimuli that share the same colored-shape are not significantly different.

Figure 4-12: Set 4 LEV rating vs. $L_J$. Each data point is labeled for its crossover time. Correlation is much lower than in previous sets ($R^2 = 0.73$) because the four data points up to the 80 ms crossover point have identical values of $L_J$, but very different LEV ratings.

Figure 4-13: Energy grid for Hall 5, the 1200-seat shoebox hall, in the 1 kHz octave band summed from 60 ms to 200 ms. Energy is shown in dB relative to the maximum. With 0° azimuth pointing toward the stage and 0° elevation pointing straight up. The images on the left show the energy distributions over spheres, while the image on the right show the same information, but flattened onto a 2-D plot (similar to an unraveled map).

Figure 4-14: Energy grid for Shoe1200 in the 1 kHz octave band summed from 200 ms to 500 ms. Energy is shown in dB relative to the maximum. The azimuthal variation is much lower than the variation from 60 ms to 200 ms as shown in Figure 4-13.
Figure 4-15: Correlation coefficient vs. rejection angle for Set 2 in the 1 kHz octave band for both front sound rejection (blue line) and rear sound rejection (orange line). The maximum $R^2$ values are circled for each case. The correlation for rejecting the front sound is maximized at 20°. The correlation for rejecting the rear sound is maximized at 70 degrees, which is a larger rejection angle than was found for the front energy. Additionally, rejecting the rear sound has a higher correlation coefficient than rejecting the front sound.

Figure 4-16: Correlation coefficient vs. early time cutoff for Set 2 in the 1 and 2 kHz octave bands. The correlation is maximized when a portion of the early sound field is included in the integration. The maximum $R^2$ values are circled for each of the octave bands shown.

Figure 4-17: Correlation coefficient vs. late time cutoff. Correlation increases slightly as the late time cutoff is increased, and asymptotically reaches a maximum around 400 ms.

Figure B-1: Low frequency measurement of a Brüel & Kjær OmniPower Type 4292 loudspeaker in an anechoic chamber.

Figure B-2: Directional response of Brüel & Kjær OmniPower Type 4292 loudspeaker [ OmniPower datasheet, bksv.com].

Figure B-3: Allowable deviations from omnidirectional directivity per octave band [ISO 3382-1:2009, adapted from Table 1].

Figure B-4: Three-way omnidirectional sound source components: subwoofer (left), mid-frequency dodecahedron (middle), and high-frequency dodecahedron (right). Note that the photos are not to scale relative to each other.

Figure B-5: Three-way crossover filters for the omnidirectional sound source.

Figure B-6: Free-field measurements of the subwoofer. Measurements on-axis with the drivers are in red and blue, and measurements with the source rotated 90° off-axis are shown in green and cyan.

Figure B-7: Directivity of the mid-frequency dodecahedron source rotated in azimuth.

Figure B-8: Directivity of the mid-frequency dodecahedron source rotated in elevation.

Figure B-9: 3D directivity plots of the high-frequency dodecahedron source in dB.

Figure B-10: Receiver locations in Eisenhower Auditorium for the omnidirectional source validation measurements.
Figure B-11: Early energy differences as a function of source rotation in dB. .................... 141

Figure B-12: Late energy differences as a function of source orientation in dB. .................. 141

Figure B-13: Differences in early decay time (EDT) as a function of source orientation. JNDs for each octave band are denoted by the orange lines, where the JND for EDT is 5%. .................................................................................................................. 142

Figure B-14: Differences in reverberation time (T30) as a function of source orientation. JNDs for each octave band are denoted by the orange lines, where the JND for T30 is 5%. .................................................................................................................. 143

Figure B-15: Differences in clarity index (C80) as a function of source orientation. JNDs for each octave band are denoted by the orange lines, where the JND for C80 is 1 dB. .................................................................................................................. 143

Figure B-16: Differences in early lateral energy fraction (J_{LF}) as a function of source orientation. JNDs for each octave band are denoted by the orange lines, where the JND for J_{LF} is 0.05. .................................................................................................................. 144

Figure B-17: Differences in late lateral energy level (L_{L}) as a function of source orientation. JNDs for each octave band are denoted by the orange lines (the JND of L_{L} is assumed to be 1 dB, the JND for strength, since the JND for L_{L} is not known). .... 144

Figure B-18: Three-way omnidirectional source in the stacked configuration. ................. 145

Figure B-19: Three-way omnidirectional source in the coincident configuration, three separate measurements. .................................................................................................................. 145

Figure B-20: Difference in early energy between the stacked configuration and coincident configuration. .................................................................................................................. 146

Figure B-21: Difference in late energy between the stacked configuration and coincident configuration .................................................................................................................. 147

Figure B-22: Differences in early decay time (EDT) for the stacked vs. coincident configurations. JNDs for each octave band are denoted by the orange lines, where the JND for EDT is 5%. .................................................................................................................. 148

Figure B-23: Differences in reverberation time (T30) for the stacked vs. coincident configurations. JNDs for each octave band are denoted by the orange lines, where the JND for T30 is 5%. .................................................................................................................. 148
Figure B-24: Differences in clarity index (C80) for the stacked vs. coincident configurations. JNDs for each octave band are denoted by the orange lines, where the JND for C80 is 1 dB.

Figure B-25: Differences in late lateral energy level (L_) for the stacked vs. coincident configurations. JNDs for each octave band are denoted by the orange lines (the JND of L_ is assumed to be 1 dB, the JND for strength, since the JND for L_ is not known).
List of Tables

Table 3-1: Correlation coefficients for different metrics as LEV predictors for Set 1............74

Table 4-1: Details about the seven halls measured as a part of this study. .......................97

Table 4-2: Approximate dimensions the seven halls measured as a part of this study. ........97

Table 4-3: Correlation coefficients between the new metric and mean LEV of the four test sets. All correlation coefficients shown are significant at p < 0.05 and the non-significant regressions are denoted by “NS”. .................................................................120

Table 4-4: Correlation coefficients between LEV rating and \( L_i \) and the correlation coefficients between LEV rating and the new metric. All correlation coefficients shown are significant at p < 0.05 and the non-significant regressions are denoted by “NS”. .............................................................................................................122
Acknowledgements

I would like to thank my thesis committee members: Drs. Sparrow, Russell, Doherty, and Rabinowitz for all of their comments, suggestions, feedback, and discussions over the last several years. I would also like to thank Chris Ickler for his helpful discussions on listener envelopment. Thanks to the professors and staff in the Graduate Program in Acoustics for providing an excellent acoustics education. Thanks to the students who have contributed to this project, including Matthew Neal, Carol Tadros, and Colton Snell. Huge shout out to Matthew Neal, who has been working on this project since the beginning and was instrumental in the construction of the loudspeaker array, conducting measurements, creating GUIs, and more. I’m excited to see his progress as he continues this work for his Ph.D.

My research involved a plethora of measurements in different halls, and I couldn’t have made them without the help from my friends. Thank you, Matthew Neal, Martin Lawless, Rachael Romond, Will Doebler, Peter Moriarty, Matthew Blevins, Laura Brill, Hyun Hong, Joonhee Lee, and Zhao Ellen Peng. The measurements would also not be possible without access to the halls, provided by Tom Hesketh, Ed Hurd, Chris Ball, Johanna Kodlick, Vonny Boarts, John Coffelt, and Jack and Carolyn Zybura.

I want to express my gratitude to my friends and family both in State College and elsewhere, especially my parents and my wife Barbara. I dragged Barbara all the way to State College from Massachusetts for a few years, and I could not have done this without her love, support, and encouragement. To all SPRALites and Research Westians, it’s been great getting to know you all both inside and outside of work.

Finally, I would like to express my greatest thanks to my advisor and chair of my committee, Dr. Michelle Vigeant. I first worked with her as an undergraduate student at the University of Hartford nearly 10 years ago. She encouraged me to pursue graduate studies, which influenced my decision to take distance education coursework in the Graduate Program in Acoustics while I was working full time. When she came to Penn State, she contacted me about this project and asked if I had ever thought about pursuing a Ph.D., which was something I didn’t think I was
capable of at the time. I took a chance when I left my job to come to Penn State, but it worked out well and I learned a ton about acoustics and conducting research. I would like to thank Michelle for believing in me, for her mentorship, for her dedication in meeting with students, and for her pursuit of perfection in conference presentations and publications.

The work presented in this dissertation was sponsored by the National Science Foundation (NSF) award #1302741. Approval for human subjects testing was obtained from Penn State’s Institutional Review Board (IRB #41733).
Chapter 1: Introduction

An important aspect of the overall impression of a concert hall is the spatial impression of the hall, which includes listener envelopment (LEV), the sense of being fully immersed in the sound field [1] [2]. Other important attributes include reverberance, bassiness, proximity, definition, and clarity [2]. The spatial impression of concert halls and LEV have been studied for decades, and objective measures have been proposed to predict the perception of LEV. The purpose of this work was to investigate LEV using state-of-the-art measurement techniques utilizing a compact spherical microphone array and subjective studies using 3D reproduction of the sound fields via higher order Ambisonics. The results from the objective analysis of the sound fields and subjective listening tests were used to inform the development of a new metric to predict LEV based on a spherical harmonic decomposition of the sound field.

Utilizing measurements obtained with a spherical microphone array has several advantages over conventional microphones that are commonly used. The impulse response (IR) measurements made using a spherical array can be used for an objective analysis of the sound field in full 3D via beamforming techniques in the spherical harmonics domain. The spherical harmonic beamforming yields a much higher spatial resolution than conventional measurement methods, which primarily use microphones with a first-order dipole or cardioid type pattern. The IRs can also be processed using Ambisonics for subjective listening tests with 3D reproductions of the sound fields over a loudspeaker array. Previously, LEV was studied primarily using simulated sound fields reproduced over a limited number of loudspeakers, which are less representative of the actual sound field experienced in a concert hall.

1.1 Listener Envelopment (LEV)

1.1.1 Early History of LEV

Research aimed at understanding the spatial perception in performing arts spaces initially focused on the directional dependence of early reflections. Originally, the sense of spaciousness was thought to be primarily associated with reverberation, but in the late 1960’s, it was found
that the spatial impression was heavily influenced by early reflections [3, 4, 5]. It was proposed that the spaciousness depended on the arrival direction of early reflections, and that stronger early lateral reflections were related to a quality referred to as “spatial responsiveness” [6].

Further work led to the development of the objective metric Early Lateral Energy Fraction, $J_{LF}$ (original notation was LF) [7],

$$J_{LF} = \frac{\int_{5\text{ms}}^{80\text{ms}} p_L^2(t)dt}{\int_{0\text{ms}}^{80\text{ms}} p^2(t)dt}$$

the ratio between the early lateral energy and the total early energy in the first 80 ms, which was found to be correlated with the subjective level of spatial impression [7]. A second metric that has been shown to correlate with spatial impression is the interaural cross correlation coefficient ($\text{IACC}_{\text{early}}$), which is obtained from the cross-correlation of the left and right ears of a binaural IR [8]:

$$\text{IACC} = \left| \frac{\int_{t_1}^{t_2} p_{\text{left}}(t)p_{\text{right}}(t+\tau)dt}{\sqrt{\int_{t_1}^{t_2} p_{\text{left}}^2(t)dt \int_{t_1}^{t_2} p_{\text{right}}^2(t)dt}} \right|_{\text{max}}$$

for $-1\text{ ms} < \tau < 1\text{ ms}$ (1-2)

where $t_1$ is 0 and $t_2$ is 80 ms for $\text{IACC}_{\text{early}}$.

Later work in spaciousness proposed that the spatial impression of a hall contains two distinct perceptions: apparent source width (ASW), which is the sense of how wide or narrow the sound image appears to a listener, and LEV, the sense of being immersed in and surrounded by the sound field [9]. ASW has since been shown to be related to early lateral reflections, which can be predicted using IACC and $J_{LF}$, while LEV has been shown to be related to late lateral energy [10].

Seminal work on LEV was conducted by Bradley and Soulodre in the early 1990’s [10, 11]. Using five loudspeakers distributed in the front half of the horizontal plane, shown in Figure 1-1, sound fields were generated with a small number of early reflections that were kept constant, and varied certain aspects of the late sound field: the reverberation time (T30), the early-to-late sound energy ratio (C80), and the strength of the late sound field ($G_{\text{late}}$). The angular distribution of the late sound was also varied, which was accomplished by playing the late sound either out
of a single frontal loudspeaker, three frontal loudspeakers spanning 70°, or five frontal loudspeakers spanning 180°. A subjective study showed that the parameters with the highest correlation to LEV were angular distribution and overall late level. These results were used to develop a metric to predict LEV called late lateral energy level, $L_J$ (prior notations $GLL$, $LG$, and $LG_{80}^\infty$) [10]:

$$L_J = 10 \log_{10} \left[ \frac{\int_{80ms}^{\infty} p_L^2(t) \, dt}{\int_{0}^{\infty} p_{10}^2(t) \, dt} \right] \text{[dB]},$$

where $p_L(t)$ is the room IR measured with a figure-of-eight microphone, and $p_{10}(t)$ is the IR of the sound source normalized at a distance of 10 meters away in a free field.

![Angular distribution of loudspeakers used to develop $L_J$](image)

*Figure 1-1: Angular distribution of loudspeakers used to develop $L_J$ (Fig. 1, Bradley and Soulodre, 1995 [10]).*

While a strong correlation was found between this metric and LEV, it should be noted that this study used a small number of loudspeakers spanning a limited angular area, and reducing the late sound field to the angles used in the study is an extreme corner case. These angular spans are not representative of real rooms, in which the late sound field is generally more diffuse. Additionally, the range of $L_J$ values of the stimuli was greater than 20 dB, which is a larger range than would be found in actual spaces.
1.1.2 Impact on LEV from energy other than late lateral arrivals

Although the earliest LEV work indicated that the late lateral energy is the component of the sound field with the highest correlation with envelopment, a number of studies have shown correlation between listener envelopment and non-lateral sound and/or early reflections.

One early study was conducted using simulated sound fields from seven loudspeakers distributed in the horizontal plane, and five loudspeakers raised 50° above the horizontal plane [12]. A key finding showed that adding a single front early reflection from above (i.e. a ceiling reflection), increases the spaciousness. Additionally, the results showed keeping lateral energy constant and adding energy from above the listener within 200 ms increases LEV.

A second subjective study was conducted using three loudspeakers in front of the listener spanning 90°, and three loudspeakers behind the listener spanning 90° [13]. The researchers varied the ratio of front energy and back energy, and found that increasing the energy behind a listener will increase LEV. Additionally, the results indicated that varying the ratio of front to back energy in the early reflections modifies the perception of LEV, although to a lesser extent than late energy. Although these findings indicate that the rear energy does play a role in the perception of LEV, this study was conducted without lateral loudspeakers, and the researchers note that reflections coming from behind the listener alone do not create a sense of envelopment.

A third subjective study was conducted using a similar loudspeaker arrangement to Bradley and Soulodre’s with the addition of a loudspeaker placed directly behind the listener and a loudspeaker overhead [14]. The late energy level simulated over the loudspeaker array was varied in four directions independently: lateral, frontal, overhead, and back while the early reflections were held constant. By varying the distribution of the late sound, findings showed that increasing late sound both above and behind the listener significantly increased LEV. The rate at which the LEV increases for increasing overhead and rear energy was found to be 30 to 50 percent of the rate at which LEV increases for increasing lateral energy.

Although these studies found that overhead and rear sound affect LEV, one study directly contradicts these findings and states that reflections above and behind do not significantly impact envelopment [15]. This study used simulated sound fields produced over eight
loudspeakers, with five loudspeakers in the same configuration as shown in Figure 1-1, and three loudspeakers that were raised in elevation either in front of or behind the listener depending on the test configuration. The researchers state that overhead and rear energy only have a slight impact on LEV, and that the LEV only increases substantially when the cosine-squared energy (i.e. $L_j$) increases. They also attribute the effects of the elevation on LEV to be an artifact of the simulated sound fields rather than an actual perceptual phenomenon.

The effect of the early sound on LEV was also investigated in a study using binaural stimuli in which the left ear signal was fed into the right channel and vice versa in order to increase interaural cross correlation [16]. Binaural IRs were measured in two different rooms, and the IRs from the two rooms were manipulated in three ways: cross-mixing only the early portion of the IR, cross-mixing only the late portion of the IR, and cross-mixing the entire IR. Findings in this study showed that cross-mixing the channels to increase the interaural cross correlation in only the early part of the IR decreased LEV. In one of the rooms, the impact to LEV was greater for cross-mixing only the early portion of the IR than it was for cross-mixing only the late portion of the IR. For both rooms, the impact to LEV was much greater for cross-mixing the full IR than it was for either the early part of the IR or the late part of the IR.

In terms of sound energy arriving from above or behind the listener, there is disagreement as to whether these arrivals are impactful for perceived LEV. It is a possibility that these anomalies are due to sound field simulation methods and reproduction over a limited number of loudspeakers. Therefore, LEV should be investigated using more realistic test stimuli which are representative of the actual space. Additionally, several studies note that the early sound field does impact LEV perception, while other studies assume that only the late sound field affects LEV perception and keep the early reflections constant throughout the study. The impact of the early sound field is not accounted for in most LEV metrics, and needs to be studied further to include its effects in an objective measure.

### 1.1.3 Other Proposed LEV Metrics

Currently, $L_j$ is the most commonly used objective metric to predict LEV, and is included in Annex A of the room acoustics measurement standard ISO 3382 [17]. However, several other objective metrics have been proposed to predict LEV. Objective measures have included energy fractions, such as late lateral energy fraction ($LLF$) [18]:
where \( p_L(t) \) is the room IR obtained with a figure-of-eight microphone and \( p \) is the room IR of the omnidirectional pressure. LLF is a similar metric to the aforementioned \( J_{LF} \), which has been shown to correlate with ASW. Unlike \( J_{LF} \), the findings from Ref. [18] indicate that LLF has very little variation from between halls, and thus is a poor predictor of LEV. However, LLF can be used along with late level \( G_{\text{late}} \) to calculate \( L_J \) [18]:

\[
L_J = G_{\text{late}} + 10 \log_{10} LLF [dB],
\]

and

\[
G_{\text{late}} = 10 \log_{10} \left( \frac{\int_{t_1}^{\infty} p_L^2(t) dt}{\int_0^{t_1} p_1^2(t) dt} \right) [dB].
\]

Since the late level has much more spread than LLF, Ref. [18] concludes that the dominant contribution to \( L_J \) is the late level component.

Late interaural cross correlation coefficient (IACC\(_{L,3}\)) has also been proposed as a metric [19][20], where Eqn. (1-2) is used with \( t_1 = 80 \) ms and \( t_2 = 85 \) ms. Similar to LLF, \( IACC_{L,3} \) has been shown to have a small spread over different halls and by itself is not a good predictor of LEV. Beranek has proposed an empirical formula to predict LEV objectively that is based on \( IACC_{L,3} \), strength (G), and clarity index (C80) [21]:

\[
LEV_{\text{calc}} = 0.5G - 10 \log_{10} \left( 1 + \log_{10} \frac{C_{80}}{10} \right) + 10 \log_{10} \left( 1 - IACC_{L,3} \right).
\]

Similar to the findings of LLF, it is likely that the dominant term is level in this formula.

Another metric that has been suggested is front/back energy ratio [22][23] (FBR):

\[
FBR = 10 \log \left( \frac{E_f}{E_b} \right),
\]
where $E_f$ and $E_b$ is the energy in the IR in the front half of the horizontal plane and the back half of the horizontal plane, respectively. However, this metric does not include an overall level term ($G$), or a term that takes lateral energy into account.

Another proposed metric is spatially balanced center time [24], in which center time $T_{Si}$ is calculated for each directional component $i$ and weighted by azimuthal arrival direction $\phi_i$:

$$T_{Si} = \frac{\int_0^\infty t \cdot p_i^2(t) dt}{\int_0^\infty p^2(t) dt}; \quad a_i = \frac{T_{Si} \cdot 1 + |\sin \phi_i|}{2}$$  \hspace{1cm} (1-9)

where $p_i(t)$ is the pressure beamformed in the specified direction, and $p(t)$ is the omnidirectional pressure. $SBT_s$ is then calculated by weighting the $a_i$ terms by the contributions from the other directions ($a_j$), and the sine of the angle in between contribution $i$ and contribution $j$, $\phi_{ij}$.

$$SBT_s = \sum_{i=1}^n \sum_{j=1}^n a_i \cdot a_j \cdot \sin \phi_{ij}$$  \hspace{1cm} (1-10)

1.1.4 LEV Summary

The most widely accepted metric to predict LEV, $L_i$, contains two components: a late level term, and a lateral energy fraction term. However, several studies have found that the early sound field and non-lateral energy can have an influence on the perception of LEV, which is not accounted for in the metric $L_i$. Evidence also suggests that the level term in $L_i$ dominates over the lateral fraction term.

Although several other metrics have been proposed to predict LEV, they have not been widely adopted in the architectural acoustics community, and there are only a limited number of studies that evaluate the performance of each of these metrics. Additionally, many of these metrics were designed using simulated sound fields that were reproduced over a limited number of loudspeakers (5-16), and may lack realism when compared to actual concert halls.

The previous work in LEV raises the following questions which are addressed in this work:

1. Several metrics contain an overall level term, most often integrated after 80 ms, while others neglect the dependence of LEV with overall level. How important is the overall level in predicting LEV?
2. The energy integration time for most LEV metrics begins at 80 ms. However, several studies have shown that early energy prior to 80 ms can impact the perception of LEV. Should early energy be included in a metric to predict LEV?

3. The most common spatial dependence used in LEV metrics is a dipole term, \( \cos(\phi) \) or \( \sin(\phi) \). The elevation dependence is also neglected, although some studies have shown overhead sound to impact LEV. How can the spatial dependence be taken into account in an LEV metric by increasing spatial resolution in the analysis and considering other spatial characteristics in both azimuth and elevation?

4. Many LEV subjective studies were conducted using simulated sound fields over a limited number of loudspeakers distributed only in the horizontal plane and do not necessarily represent a realistic sound field in a concert hall. Can these methods be improved upon by conducting listening tests utilizing 3D reproductions of measured sound fields?

1.2 SPHERICAL MICROPHONE ARRAYS

1.2.1 Wave equation in spherical coordinates

For spherical microphone array processing, it is necessary to use the solutions to the wave equation in spherical coordinates [25] [26]. The most general form of the linearized wave equation for sound waves in air is:

\[
\nabla^2 p - \frac{1}{c^2} \frac{\partial^2 p}{\partial t^2} = 0 . \tag{1-11}
\]

where \( p \) is the acoustic pressure, \( t \) is time, and \( c \) is the wave speed. The Laplacian operator \( \nabla^2 p \) in spherical coordinates is equal to:

\[
\nabla^2 p = \frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial p}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial p}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 p}{\partial \phi^2} , \tag{1-12}
\]

where \( r \) is the radius, \( \phi \) is the azimuth angle, and \( \theta \) is the elevation angle. Plugging Eqn. (1-12) into Eqn. (1-11) yields the wave equation in spherical coordinates:

\[
\frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial p}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial p}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 p}{\partial \phi^2} - \frac{1}{c^2} \frac{\partial^2 p}{\partial t^2} = 0 . \tag{1-13}
\]
The solutions to this partial differential equation are obtained through the technique of separation of variables, in which the solution is assumed to be the product of four separate functions:

\[ p(r, \phi, \theta, t) = R(r)\Phi(\phi)\Theta(\theta)T(t). \]  \hspace{1cm} (1-14)

Substituting this solution into the wave equation yields:

\[ \frac{1}{r^2 R} \frac{\partial}{\partial r} \left( r^2 \frac{\partial R}{\partial r} \right) + \frac{1}{r^2 \sin \theta \Theta} \frac{\partial}{\partial \theta} \left( \cos \theta \frac{\partial \Theta}{\partial \theta} \right) + \frac{1}{r^2 \Phi \sin^2 \theta} \frac{\partial^2 \Phi}{\partial \phi^2} = \frac{1}{c^2 T} \frac{\partial^2 T}{\partial t^2}. \]  \hspace{1cm} (1-15)

Since the time dependent functions are completely isolated to the right hand side of the equation and the spatial dependent functions are isolated to the left hand side of the equation, both sides of the equation must equal a constant, \(-k^2\). Taking only the right hand side of the equation yields the ordinary differential equation:

\[ \frac{1}{c^2 T} \frac{\partial^2 T}{\partial t^2} = -k^2, \]  \hspace{1cm} (1-16)

which has solutions:

\[ T(t) = Ae^{i\omega t} + Be^{-i\omega t}, \]  \hspace{1cm} (1-17)

where \(k\) is the wave number, and \(\omega = ck\) is the angular frequency. Eqn. (1-15) can be rearranged to isolate the azimuthal dependent functions to the right hand side of the equation, which must equal a constant again, \(m^2\):

\[ \sin^2 \theta \left[ \frac{1}{R} \frac{\partial}{\partial r} \left( r^2 \frac{\partial R}{\partial r} \right) + \frac{1}{\Theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial \Theta}{\partial \theta} \right) + k^2 r^2 \right] = \frac{1}{\Phi} \frac{\partial^2 \Phi}{\partial \phi^2} = m^2. \]  \hspace{1cm} (1-18)

The middle and right hand sides of the equation are now an ordinary differential equation of the same form as the time dependence:

\[ \frac{1}{\Phi} \frac{\partial^2 \Phi}{\partial \phi^2} = -m^2, \]  \hspace{1cm} (1-19)

with solutions of:
\[\Phi(\phi) = Ce^{im\phi} + De^{-im\phi}, \quad (1-20)\]

where \( m \) will be referred to as the degree of the spatial functions, and \( C \) and \( D \) are arbitrary constants. Rearranging Eqn. (1-18), the radial dependent functions are isolated to the left hand side of the equation and the elevation dependent functions are isolated to the right hand side of the equation, which both must equal a constant \( C \):

\[
\frac{1}{r} \frac{\partial}{\partial r} \left( r^2 \frac{\partial R}{\partial r} \right) + k^2 r^2 = \frac{m^2}{\sin^2 \theta} - \frac{1}{\theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial \Theta}{\partial \theta} \right) = C. \quad (1-21)
\]

Taking the right hand side of the equation and applying the transformation \( z = \cos(\theta) \), Eqn. (1-21) can be rewritten as:

\[
(1 - z^2) \frac{d^2 \Theta(z)}{dz^2} - 2z \frac{d\Theta(z)}{dz} + \left( C - \frac{m^2}{1 - z^2} \right) \Theta(z) = 0. \quad (1-22)
\]

Letting \( C = n(n - 1) \), the solutions for this ordinary differential equation are associated Legendre polynomials \( P_n^m \) of order \( n \) and degree \( m \):

\[
\Theta(z) = E P_n^m(z) = E P_n^m(\cos \theta), \quad (1-23)
\]

where \( E \) is an arbitrary constant. The left hand side of Eqn. (1-21) can be rearranged into spherical Bessel equation:

\[
\frac{\partial^2 R}{\partial r^2} + \frac{2}{r} \frac{\partial R}{\partial r} + \left( k^2 - \frac{n(n + 1)}{r^2} \right) R = 0. \quad (1-24)
\]

The solutions to this ordinary differential equation are spherical Hankel functions:

\[
R(r) = F h_n^{(1)}(kr) + G h_n^{(2)}(kr). \quad (1-25)
\]

Substituting the solutions into Eqn. (1-14) yields the complete solution to the wave equation:

\[
p(r, \phi, \theta, t) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} A_{mn} \left\{ e^{im\phi} \right\} \left\{ h_n^{(1)}(kr) \right\} P_n^m(\cos \theta) \left\{ e^{i\omega t} \right\} \left\{ e^{-im\phi} \right\} \left\{ h_n^{(2)}(kr) \right\}, \quad (1-26)
\]

where \( A_{mn} \) is a combined constant of order \( n \) and degree \( m \). For the purposes of this work, the \( e^{i\omega t} \) time convention will be used, which makes \( h_n^{(2)}(kr) \) an outward travelling wave and \( h_n^{(1)}(kr) \) an inward traveling wave. Alternatively, complex exponentials can be replaced with
real valued sin( ) and cos( ), and complex Hankel functions can be replaced with real valued Bessel functions, which is more convenient for standing waves:

\[ p(r, \phi, \theta, t) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \hat{A}_{m,n} \left\{ J_n(kr) \left\{ \cos(m\phi) \right\} \right\} \left\{ P_n^m(\cos \theta) \left\{ \cos(\omega t) \right\} \right\}. \]  (1-27)

For convenience, the angular terms can be combined into functions called spherical harmonics of order \( n \) and degree \( m \):

\[ Y^m_n(\phi, \theta) = \frac{2n + 1}{4\pi} \frac{(n - m)!}{(n + m)!} P_n^m(\cos \theta) e^{im\phi}. \]  (1-28)

The real part of \( Y^m_n(\phi, \theta) \) are shown in Figure 1-2 for orders 0 through 3.

![Figure 1-2: Plots of the real-valued spherical harmonics for orders \( n = 0, 1, 2, \) and 3.](image)

The constant term in front makes the spherical harmonics form an orthonormal basis set, with the orthogonality property:

\[ \int_{0}^{\pi} \int_{0}^{2\pi} Y^m_{n'}(\theta, \phi) Y^m_n(\theta, \phi) \sin \theta \, d\theta \, d\phi = \delta_{mm'} \delta_{nn'}, \]  (1-29)

allowing the spherical harmonic functions to be used for a spherical Fourier transform. Using spherical harmonics, the solution to the wave equation becomes:
\[ p(r, \phi, \theta, t) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \left[ A_{mn} h^{(1)}_{n}(kr) + B_{mn} h^{(2)}_{n}(kr) \right] Y^m_n(\phi, \theta) e^{i\omega t}, \quad (1-30) \]

Or alternatively replacing with spherical Hankel functions with spherical Bessel functions:

\[ p(r, \phi, \theta, t) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \left[ C_{mn} j_{n}(kr) + D_{mn} y_{n}(kr) \right] Y^m_n(\phi, \theta) e^{i\omega t}. \quad (1-31) \]

### 1.2.2 Spherical Fourier Transform of Microphone Array Signals

Any function of \((\theta, \phi)\) can be represented as an infinite sum of spherical harmonics with a weighting coefficient:

\[ f(\theta, \phi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} f_{nm} Y^m_n(\theta, \phi). \quad (1-32) \]

By exploiting the orthogonality of the spherical harmonic functions (Eqn. (1-29)), the coefficients can be extracted by:

\[ f_{nm} = \frac{2\pi}{Q} \int_{0}^{\pi} \int_{0}^{2\pi} f(\theta, \phi) Y^m_n(\theta, \phi) \sin \theta \, d\theta \, d\phi. \quad (1-33) \]

Using a uniform sampling scheme that distributes the sampled points equally in solid angle on the surface of the sphere, such as arrays in which the sample points are faces or vertices of regular polyhedra, the discrete orthogonality condition becomes:

\[ \frac{4\pi}{Q} \sum_{q=1}^{Q} Y^m_{n'}(\theta_q, \phi_q) Y^m_n(\theta_q, \phi_q) = \delta_{mm'} \delta_{nn'}, \quad (1-34) \]

where \(Q\) are the total number of sample points, and \((\theta_q, \phi_q)\) are the sample locations on the sphere. The orthogonality condition can be used to develop a discrete spherical Fourier transform to extract the spherical harmonic coefficients:

\[ f_{nm} = \frac{4\pi}{Q} \sum_{q=1}^{Q} f(\theta_q, \phi_q) Y^m_n(\theta_q, \phi_q), \quad (1-35) \]
One consequence of sampling is that the spherical Fourier transform becomes order-limited to a maximum order of $N$. The number of samples required for a given order $N$ is $Q = (N + 1)^2$. A second consequence of sampling is that spatial aliasing is introduced at high frequencies and high orders, in which the spacing of the microphones are large and the sample points do not satisfy the Nyquist sampling criterion.

Applying Eqn. (1-35) to pressure signals measured from a spherical microphone array, the discrete spherical Fourier transform becomes:

$$
\tilde{P}_{nm}(\omega) = \frac{4\pi}{Q} \sum_{q=1}^{Q} P_q(\omega) Y_{nm}^*(\theta_q, \phi_q),
$$  \hspace{1cm} (1-36)

where $P_q(\omega)$ is the complex pressure in the frequency domain measured at microphone $q$, and $\tilde{P}_{nm}(\omega)$ are the spatial Fourier coefficients in the spherical harmonics domain as a function of frequency.

1.2.3 Scattering off of a rigid sphere

The microphone array can be modeled as a rigid sphere with radius $a$. In order to correct for the placement of the microphones on the sphere, the scattered sound field must be taken into account [25] [26]. Consider an incident plane wave that can be expanded into an infinite sum of Bessel functions $j_n(kr)$ and spherical harmonics $Y_n^m(\theta, \phi)$:

$$
p_i(r, \theta, \phi, t) = P_0 4\pi i^n \sum_{n=0}^{\infty} j_n(kr) \sum_{m=-n}^{n} Y_n^m(\theta, \phi) Y_n^m(\theta_i, \phi_i) e^{i\omega t},
$$  \hspace{1cm} (1-37)

The assumed form of the scattered wave is Eqn. (1-30):

$$
p_s(r, \theta, \phi, t) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} c_{nm} h_n^{(2)}(kr) Y_n^m(\theta, \phi) e^{i\omega t},
$$  \hspace{1cm} (1-38)

where $c_{nm}$ is a constant for each order $n$ and degree $m$. The coefficient of $h_n^{(1)}(kr)$ is assumed to be zero since the scattered wave will be outward travelling. The $c_{nm}$ coefficients are calculated by applying the boundary condition that the radial particle velocity $u(r, \theta, \phi, t)$ is equal to zero on the surface of the rigid sphere:
Here $u_i$ is the incident contribution and $u_s$ is the scattered contribution to the total field $u$. Using Euler’s equation, the boundary condition can be rewritten substituting the particle velocity with the partial derivative of pressures with respect to $r$:

$$\frac{\partial}{\partial r} p_i|_{r=a} = -\frac{\partial}{\partial r} p_i|_{r=a}. \quad (1-40)$$

Substituting Eqns. (1-37) and (1-38) into Eqn. (1-40) yields:

$$P_0 4\pi i^n \sum_{n=0}^{\infty} j'_n (ka) \sum_{m=-n}^{n} Y_n^m (\theta, \phi) Y_n^m*(\theta_i, \phi_i) e^{i\omega t}$$

$$= \sum_{n=0}^{\infty} \sum_{m=-n}^{n} c_{nm} h^{(2)}_n (ka) Y_n^m (\theta, \phi) e^{i\omega t}, \quad (1-41)$$

Equating each term in the summation gives:

$$P_0 4\pi i^n j'_n (ka) Y_n^m (\theta, \phi) Y_n^m*(\theta_i, \phi_i) e^{i\omega t}$$

$$= c_{nm} h^{(2)}_n (ka) Y_n^m (\theta, \phi) e^{i\omega t}, \quad (1-42)$$

which can be rearranged to solve for $c_{nm}$:

$$c_{nm} = \frac{P_0 4\pi i^n j'_n (ka) Y_n^m*(\theta_i, \phi_i)}{h^{(2)}_n (ka)}. \quad (1-43)$$

The total expression for the scattered field is therefore:

$$p_s(r, \theta, \phi, t)$$

$$= P_0 4\pi i^n \sum_{n=0}^{\infty} \frac{j'_n (ka) h^{(2)}_n (kr)}{h^{(2)}_n (ka)} \sum_{m=-n}^{n} Y_n^m (\theta, \phi) Y_n^m*(\theta_i, \phi_i) e^{i\omega t}. \quad (1-44)$$

The total pressure on the surface of the sphere, which is the pressure seen by the microphones on the spherical array, is:
\[ p_i + p_s = P_0 4\pi i^n \sum_{n=0}^{\infty} b_n \sum_{m=-n}^{n} Y_n^m(\theta, \phi) Y_n^m(\theta_i, \phi_i) e^{i\omega t} \] (1-45)

where \( b_n \), sometimes referred to as plane wave modal coefficients, are:

\[ b_n = j_n(ka) + \frac{j_n^{(2)}(ka)}{h_n^{(2)}(ka)} h_n^{(2)}(ka), \] (1-46)

as shown in Figure 1-3 below.

For the spherical microphone array, the scattered field is beneficial because the scattering term fills in the holes in the frequency response of the array that correspond to the zeros in the \( j_n(ka) \) Bessel functions.

In order to equalize the frequency response of the spherical harmonics components as measured on the surface of the sphere, a factor of \( 1/b_n(ka) \) must be applied to the measurements. This process is known as radial filtering.

1.2.4 Beamforming

The spherical harmonic components can be weighted and summed to form directional beams. The purpose of beamforming, in the context of this research, is to isolate the sound energy that is being received in a particular direction. Axisymmetric beamforming is a convenient and
efficient beamforming method in which the \( m = 0 \) components (i.e. the zero-degree components) that are functions of \( P_n(\cos \theta) \) are weighted and summed to form the desired beampattern, and this pattern is steered to the desired look direction of the beam (the direction in which the main lobe of the beam pattern is oriented).

The spherical harmonics addition theorem states:

\[
\sum_{m=-n}^{n} Y_n^m(\theta_l, \phi_l) Y_n^{m*}(\theta, \phi) = \frac{2n + 1}{4\pi} P_n(\cos \Psi),
\]

where \( \Psi \) is the angle between \((\theta, \phi)\) and the desired look direction \((\theta_l, \phi_l)\):

\[
\cos \Psi = \cos \theta \cos \theta_l + \cos(\theta - \theta_l) \sin \theta \sin \theta_l.
\]

In other words, the zero-degree spherical harmonic of order \( n \) can be oriented to point in the desired look direction \((\theta_l, \phi_l)\) by multiplying each degree component by that component’s spherical harmonic evaluated in the look direction, and summing over spherical harmonic degree \( m \). Once the zero degree components are steered in the proper direction, they can be weighted and summed to form the desired beam pattern:

\[
y(\theta, \phi) = \sum_{n=0}^{N} c_n \sum_{m=-n}^{n} Y_n^m(\theta_l, \phi_l) Y_n^{m*}(\theta, \phi),
\]

where \( c_n \) are order dependent weights. By setting \( c_n = 1 \), the beam shape is a plane wave truncated to order \( N \). For this reason, setting the weights to unity is often referred to as a “plane wave decomposition” (PWD) beamformer. An example of the beam pattern for a truncated plane wave of order \( N = 3 \) is shown in Figure 1-4.
Applying this beamforming technique to the microphone signals that have been transformed into the spherical harmonics domain can be accomplished by:

\[
y_c(f, \theta_l, \phi_l) = \sum_{n=0}^{N} c_n \sum_{m=-n}^{n} \tilde{P}_{nm}(f) Y_n^m(\theta_l, \phi_l). \tag{1-50}
\]

The overall beamforming system can be composed utilizing all of the components introduced in Sections 1.2.2 through 1.2.4. A block diagram of the beamforming system is shown in Figure 1-5.

Figure 1-5: Block diagram for the overall beamforming system (adapted from Fig. 5.3, Fundamentals of Spherical Array Processing, Rafaely, 2015 [25]).

1.3 AMBISONICS

Third-order Ambisonics (more generally referred to here as Ambisonics) was utilized in this work to generate spatial reproductions of the measured data. Ambisonics is a spatial audio playback system originally developed by Gerzon in the 1970’s as a method to reproduce sound fields represented in the spherical harmonics domain [27]. Ambisonics initially only used the zeroth and first order spherical harmonic components, and has since been extended to higher orders [28].
Ambisonics offers a convenient method of reproducing recordings obtained with a compact spherical microphone array since the processing is done in the spherical harmonics domain [29].

### 1.3.1 Spherical Harmonics Format

For the Ambisonic reproduction, this work uses a form of the spherical harmonics that employs real-valued trigonometric functions, which are more convenient for audio signals, based on the ambiX format convention [30]:

\[ \hat{Y}^n_m(\alpha, \phi) = \sqrt{\frac{(2 - \delta_{m0}) (n - |m|)!}{4\pi (n + |m|)!}} P^n_m(\sin \alpha) \left\{ \begin{array}{ll} \sin(|m\phi|) & \text{if } m < 0 \\ \cos(|m\phi|) & \text{if } m \geq 0 \end{array} \right. \]  

(1-51)

where \( \alpha = \frac{\pi}{2} - \theta \) is the elevation angle relative to the horizontal plane, and \( \delta \) is the Kronecker delta. Although not standard notation, the hat is used here to differentiate between the spherical harmonics. This format is preferred over standard spherical harmonics for several reasons:

1. The constant term in front ensures that the value of each component components will not exceed the value of the 0\(^{th}\) order component, which helps to prevent audio clipping. However, this change means that the functions are no longer orthonormal (although they are still orthogonal).

2. The elevation angle is more convenient for audio reproductions, since \( \alpha = 0^\circ \) lies on the horizontal plane.

3. The real-valued trigonometric functions are more convenient to use with real-valued time domain signals.

### 1.3.2 Ambisonic Decoding

Time domain microphone signals can be encoded (i.e. transformed) into Ambisonic (i.e. spherical harmonic) signals using spherical harmonics as described in Section 1.2:

\[ \hat{p}^m_n(t) = \frac{4\pi}{Q} \sum_{q=1}^{Q} p_q(t) \hat{Y}^n_m(\alpha_q, \phi_q) \]  

(1-52)

Radial filtering needs to be applied to \( \hat{p}^m_n(t) \) to equalize each component’s signal. The Ambisonic signals are then decoded into loudspeaker signals, which is the inverse of the
encoding scenario. The loudspeakers sample each spherical harmonic in space, and the proper gains need to be applied to recreate the desired Ambisonic signals:

\[
\hat{p}_{nm}(t) = \sum_{l=1}^{L} g_l(t) \hat{Y}_{m}^{n}(\alpha_l, \phi_l),
\]

(1-53)

where \(g_l(t)\) are the individual loudspeaker signals, \((\alpha_l, \phi_l)\) are the elevation and azimuth angles of loudspeaker \(l\), and \(L\) is the total number of loudspeakers. This equation may be written in matrix form:

\[
\hat{p}_{nm} = \hat{Y} \mathbf{g},
\]

(1-54)

where

\[
\hat{p}_{nm} = \begin{bmatrix} \hat{p}_{00}, \hat{p}_{1(-1)}, \hat{p}_{10}, \hat{p}_{11}, \ldots, \hat{p}_{NN} \end{bmatrix}^T,
\]

(1-55)

\[
\mathbf{g} = \begin{bmatrix} g_1(t), g_2(t), \ldots, g_L(t) \end{bmatrix}^T,
\]

(1-56)

and

\[
\hat{Y} = \begin{bmatrix} \hat{Y}_0^0(\alpha_1, \phi_1) & \hat{Y}_1^{-1}(\alpha_1, \phi_1) & \ldots & \hat{Y}_N^N(\alpha_1, \phi_1) \\
\hat{Y}_0^0(\alpha_2, \phi_2) & \hat{Y}_1^{-1}(\alpha_2, \phi_2) & \ldots & \hat{Y}_N^N(\alpha_2, \phi_2) \\
\vdots & \vdots & \ddots & \vdots \\
\hat{Y}_0^0(\alpha_L, \phi_L) & \hat{Y}_1^{-1}(\alpha_L, \phi_L) & \ldots & \hat{Y}_N^N(\alpha_L, \phi_L) \end{bmatrix}.
\]

(1-57)

To solve for the loudspeaker driving signals, \(\mathbf{g}\), a least-square solution is obtained by taking the pseudo-inverse of spherical harmonic matrix \(\hat{Y}\):

\[
\mathbf{g} = \hat{Y}^\dagger \hat{p}_{nm} = \mathbf{D} \hat{p}_{nm},
\]

(1-58)

where \(\mathbf{D} = \hat{Y}^\dagger\) is known as the basic decoder matrix, since it decodes Ambisonic (i.e. spherical harmonic) signals into loudspeaker driving signals. This method to design an Ambisonic decoder is known as the mode-matching or pseudoinverse method [31] [32, 33].

1.3.3 Max-RE Decoding

Gerzon introduced two metrics that can be used to evaluate Ambisonic reproduction systems [34]. The first is the velocity vector:
\[ r_V = Re \left( \frac{\sum_{l=1}^{L} G_l \hat{u}_l}{\sum_{l=1}^{L} G_l} \right), \]  

(1-59)

and the second is the energy vector:

\[ r_E = \frac{\sum_{l=1}^{L} G_l G_l^* \hat{u}_l}{\sum_{l=1}^{L} G_l G_l^*}, \]  

(1-60)

where \( G_l \) are the loudspeaker gains, which can in general be complex, and \( \hat{u}_l \) is a unit vector pointed in the direction from the listener to the loudspeaker. For optimal Ambisonic reproduction, the magnitudes of the velocity and energy vectors need to be constant over frequency and point in the same direction. At low frequencies, it is important that the magnitude of the velocity vector to be close to 1 for all incident angles, which will ensure accurate reproduction of the interaural time difference (ITD) localization cues. At high frequencies, the energy vector must be maximized for as many angles as possible to ensure accurate reproduction of the interaural level difference (ILD) localization cues. The basic decoding matrix given in Eqn. (1-58) is the decoder solution that maximizes the velocity vector, and is therefore appropriate as a low frequency decoder.

To maximize the energy vector at high frequencies, an additional order-dependent gain can be applied to each spherical harmonic component. The effect of applying the order dependent gain, shown in Figure 1-6, is that the order-limited plane wave that is received at the listening location has reduced side lobes, minimizing energy coming from directions other than the intended arrival direction. The reduced side lobes come at the expense of a wider main lobe. This scheme is referred to as Max-\( r_E \) decoding [28] [35].
Figure 1-6: Energy received at the listening location from a plane wave produced using 3rd order Ambisonics for basic decoding (left) and Max-\(r_E\) decoding (right).

For 3rd order Ambisonics, the order-dependent Max-\(r_E\) gains are [32]:

\[
c_n = [1, 0.861, 0.612, 0.305]
\]  

(1-61)

### 1.3.4 Nearfield Compensation

The sound radiated from a loudspeaker at low frequencies can be modeled as a point source, which can be expanded into spherical harmonics:

\[
\frac{P_0 e^{i(\omega t - kr_s)}}{r_s} = P_0 4\pi (-i) k \sum_{n=0}^{\infty} h_n^{(2)}(kr_s) j_n(kr) \sum_{m=-n}^{n} Y_n^m(\theta, \phi) Y_n^m(\theta_i, \phi_i) e^{i\omega t},
\]  

(1-62)

where \(r_s\) is the distance of the source to the listening position. Dividing Eqn. (1-62) by Eqn. (1-37) yields the order-dependent pressure gain between point source radiation and plane wave radiation:

\[
\frac{p_{\text{point}}(kr)}{p_{\text{plane}}(kr)} = i^{-(n-1)} k h_n^{(2)}(kr_s) j_n.
\]  

(1-63)

The effect of the point source radiation is that for low frequencies where \(kr_s < n\), the pressure rises as frequency decreases at 6\(n\) dB/octave. Therefore, nearfield compensation filters need to be applied that invert the response of Eqn. (1-63).
1.3.5 AURAS Facility

The Ambisonics reproduction in this work was conducted in the Auralization and Reproduction of Acoustic Sound-fields (AURAS) facility at Penn State University, shown in Figure 1-7a. The facility includes a loudspeaker array consisting of 30 two-way sealed-box loudspeakers. The loudspeakers feature a 4” (~10 cm) mid-bass driver and a 1” (~2.5 cm) fabric dome tweeter that are passively crossed over at 1.8 kHz. The loudspeakers are individually equalized from approximately 60 Hz to 20 kHz to account for magnitude and phase differences in the frequency response of each loudspeaker. Details on the design and construction of the loudspeaker array can be found in Ref. [36].

1.3.5.1 Layout

The 30 loudspeakers in the AURAS array are arranged in a nearly-spherical distribution as shown in Figure 1-7b. Twenty-eight of the loudspeakers are located in 3 rings: 8 loudspeakers located at $\alpha = -30^\circ$, 12 loudspeakers located at $\alpha = 0^\circ$, and 8 loudspeakers located at $\alpha = +30^\circ$. In each ring, the loudspeakers are distributed equally azimuthally with a loudspeaker located at $\phi = 0^\circ$ directly in front of the listener. The average distance from the loudspeakers to the center of the array is $r = 1.3$ m. The remaining two loudspeakers are placed overhead at $\alpha = +60^\circ$, $\phi = \pm 90^\circ$, and $r = 0.57$ m. This loudspeaker layout maintains spherical harmonic orthogonality up to order $N = 3$.

1.3.5.2 AURAS Ambisonic Decoder

The Ambisonics Decoder Toolbox [32, 33] was utilized to generate a VST plugin to accomplish Ambisonic decoding over the loudspeaker array. This decoder was designed using the pseudoinverse method as described in Section 1.3.2. The decoder is a two-band decoder, with
basic decoding at low frequencies and Max-\(r_E\) decoding at high frequencies, as described in Section 1.3.3, implemented with a phase-matched crossover filter at 400 Hz. To account for the distance of the loudspeakers from the center of the array, the decoder also has level and time delay compensation to account for the \(1/r\) pressure dependence and propagation delay, respectively. Additionally, nearfield compensation filters are applied in the decoder, as described in Section 1.3.4.

**1.3.5.3 Ambisonic Performance**

The performance of the Ambisonic decoder was evaluated by looking at the magnitude and direction of the \(r_V\) and \(r_E\) vectors in the Ambisonics Decoder Toolbox [32, 33], where the \(r_V\) and \(r_E\) vectors are used to evaluate the low and high frequency performance, respectively, of an Ambisonic setup. The closer the magnitude of the individual vectors for a given setup are to the possible maximum value for each vector, the better the performance of the system. Additionally, the direction of the vectors should point in the intended direction, and the direction of the \(r_V\) and \(r_E\) vectors should be the same.

As shown in Figure 1-8, the \(r_V\) vector of the AURAS decoder has a magnitude of unity, which is the maximum value, over the entire sphere. Additionally, the angular error is zero over the entire sphere. These results indicate that the Ambisonic array should have excellent low frequency performance over the entire 3D space.

For a 3rd order Ambisonics system, the maximum value of the magnitude of the \(r_E\) vector is 0.86 [32]. As shown in Figure 1-9, the magnitude of the \(r_E\) vector for the AURAS decoder approaches this value over most of the sphere. However, the magnitude of \(r_E\) is low in regions of the sphere where there is no loudspeaker coverage, primarily below \(\alpha = -30^\circ\) where the magnitude of \(r_E\) falls to a minimum value of 0.14. In the top hemisphere, the minimum value of \(r_E\) is 0.7 at \(\alpha = 60^\circ\) and \(\phi = 0^\circ\) where no loudspeaker is present. Similarly, the angular error of \(r_E\) is low over most of the sphere, aside from regions where there are no loudspeakers. These results suggest that the system should perform well at high frequencies, provided there are loudspeakers present for a given direction. Further validation of the Ambisonic reproductions in the AURAS facility are detailed in Chapter 3.
Figure 1-8: Magnitude, direction, and angular error of the $r_v$ vector, plotted in the Ambisonics Decoder Toolbox.
1.4 DISSERTATION OUTLINE

The remainder of this dissertation is organized as follows:

Chapter 2 will detail the spherical microphone array measurement setup and a validation study where measured room acoustic metrics obtained with the spherical microphone array are compared to measurements made with conventional microphones. Note that this chapter is a reproduction of D. A. Dick and M. C. Vigeant, "A comparison of measured room acoustics metrics using a spherical microphone array and conventional methods," *Appl. Acoust.*, vol. 107, pp. 34-45, 2016 [37]. Chapter 3 will discuss spatial room IR measurements obtained in a concert hall, two subjective studies using the room IRs, an evaluation of several metrics used to predict LEV, and trends found from LEV ratings compared with an objective analysis of the sound field. Chapter 4 will discuss spatial room IR measurements obtained in seven different halls, a subjective study comparing the LEV ratings of the halls, and will outline the development of a
new metric to predict LEV based on spherical harmonics. Chapter 5 will summarize the conclusions and recommend future work.
Chapter 2: A Comparison of Measured Room Acoustics Metrics Using a Spherical Microphone Array and Conventional Methods

This chapter will detail the spherical microphone array measurement setup and a validation study where measured room acoustic metrics obtained with the spherical microphone array are compared to measurements made with conventional microphones. The text in this chapter is a reproduction of an article in Applied Acoustics published in June 2016. This chapter differs from the published article in formatting and minor wording changes suggested by the committee for clarity.


DOI: https://doi.org/10.1016/j.apacoust.2016.01.008

\(^1\) https://www.elsevier.com/about/our-business/policies/copyright/permissions [last accessed 11/14/17]: “Authors can include their articles in full or in part in a thesis or dissertation for non-commercial purposes.”
A Comparison of Measured Room Acoustics Metrics Using a Spherical Microphone Array and Conventional Methods

ABSTRACT

The traditional microphone configuration used to measure room impulse responses (IRs) according to ISO 3382:2009 is an omnidirectional and figure-8 microphone pair. IRs measurements were taken in a 2500-seat auditorium to determine how the results from a spherical microphone array (an mh acoustics Eigenmike-em32) compare to those from the traditional microphone setup (a Brüel & Kjær Type-4192 omnidirectional microphone and a Sennheiser MKH30 figure-8 microphone). Measurements were obtained at six receiver locations, with three repetitions each in order to first evaluate repeatability. The metrics considered in this study were: reverberation time (T30), early decay time (EDT), clarity index (C80), strength (G), lateral energy fraction (JLF) and late lateral energy level (L). Before calculating these quantities, the IRs were filtered to equalize the frequency response of the microphones and sound source. For the spherical array measurements, the omnidirectional (monopole) and figure-8 (dipole) patterns were extracted using beamforming. In terms of repeatability, the average standard deviation of the three measurements at each receiver location averaged across all metrics, receivers, and octave bands was found to be 0.01 just noticeable differences (JNDs). The analysis comparing the measurements from the two microphone configurations yielded differences which were less than 1 JND for the majority of metrics, with a few exceptions of EDT and C80 slightly above 1 JND. Based on this case study, these results indicate that spherical microphone arrays can be used to obtain valid room IR measurements, which will allow for the development of new metrics utilizing the higher spatial resolution made possible with spherical arrays.

2.1 INTRODUCTION

Spherical microphone arrays contain a number of microphones arranged on the surface of a compact sphere and can be used to obtain spatial information about sound fields. The spherical configuration of the array enables a convenient way to beamform directional patterns in any direction in 3D space using a spatial Fourier transform and processing the signals in the spherical
harmonics domain [25] [38]. In recent years, spherical microphone arrays have begun to be utilized in room acoustics applications to analyze the directional properties of reverberant spaces [3-6]. Room impulse responses (IRs) measured with spherical arrays have been analyzed to determine the direction of arrival of early reflections in rooms [39, 40, 41]. A recent study also evaluated IR measurements obtained in performing arts spaces using a 16-channel spherical microphone array by beamforming the IRs in the azimuthal plane and comparing different audience receiver positions [42].

Previous room acoustics studies involving spherical microphone arrays have not included analyses of the IRs to calculate established room acoustics metrics as defined in Annex A of ISO 3382 [43]. These metrics require measurements made using a pair of microphones, one with an omnidirectional directivity pattern and a second one with a figure-of-eight (figure-8) directivity pattern. Alternatively, these directivity patterns can be obtained from spherical microphone array measurements by extracting the zeroth (monopole) and first order (dipole) spherical harmonic components, respectively. Before this analysis can be done, however, room acoustics metrics using spherical microphone arrays must be verified against traditional methods in order to gain confidence that the measurements are consistent. This research is especially necessary due to the fact that previous work has shown a large variation in measured parameters made between different microphone types and with different measurement teams [20-25]. Additionally, this comparison is necessary since spherical microphone arrays are generally larger than conventional measurement microphones, and therefore may alter the sound field if the sound wave that is scattered from the array reflects off of nearby objects and returns to the microphone array [25].

Obtaining room acoustics metrics with spherical microphone arrays may offer some advantages compared to measurements made with conventional microphones. Spatial measures are typically obtained using a figure-8 microphone. Commercially available figure-8 microphones are not laboratory-grade and may not have ideal directivity, frequency response, or linearity; whereas spherical array microphones are typically constructed using laboratory-grade microphone capsules. Spherical microphone arrays also enable the researcher to rotate the figure-8 pattern in post-processing to perfectly align the pattern to the source, which could reduce measurement uncertainty. Finally, current spherical microphone array technology
enables beamforming utilizing spherical harmonics up to third- or fourth-order, which can be used to create new room acoustics metrics with a much higher spatial resolution than the traditional first-order dipole.

The purpose of this case study was to compare measurements taken in accordance with the ISO 3382 standard using a traditional omnidirectional and figure-8 microphone pair with measurements taken using a spherical microphone array. This comparison is required in order to gain confidence that room acoustics measurements made with a spherical microphone array can be directly compared to measurements made with traditional methods. Once this verification is complete, new metrics with higher spatial resolution can be developed.

2.2 ROOM ACOUSTICS METRICS

The metrics that were evaluated in this study are defined in ISO 3382 and accompanying Annex A. The omnidirectional measures are reverberation time (T30), measured from a 30 dB decay from the Schroeder backwards integrated curve; early decay time (EDT), measured from the slope of the first 10 dB decay of the Schroeder backwards integrated curve; clarity index (C80), the ratio of the early sound in the first 80 ms to the late sound; and strength (G), the energy in the room IR normalized to the level of the sound source measured at a distance of 10 m in a free field. In addition to the commonly used omnidirectional measures, metrics used to predict the spatial impression of a room are included in Annex A of ISO 3382. Spatial impression is one characteristic that has been shown to be related to overall room impression [44, 1, 2]. Previous research proposed that spatial impression should be formally divided into two distinct components [11]: the apparent source width (ASW) as being associated with the early lateral reflections, and listener envelopment (LEV), which is related to late lateral reflections [45]. A number of objective measures have been proposed to predict both ASW and LEV that utilize either directional microphones or a binaural head [19]. The two spatial metrics that have gained the largest acceptance in the architectural acoustics community are early lateral energy fraction ($J_{LF}$, previously $LF$) [46], which is used to predict ASW, and late lateral energy level ($L_J$, previously $GLL$, $L_G$, and $L_{G80}$) [45], which is used to predict LEV. Both of these metrics are included in ISO 3382 Annex A and were evaluated as part of this study. $J_{LF}$ is the ratio of early lateral energy to total early energy:
where $p_f(t)$ is the IR measured with a figure-8 microphone, and $p_o(t)$ is the IR measured with an omnidirectional microphone. $L_J$ is the ratio of the late lateral energy to the normalized source energy:

$$L_J = 10 \log \left[ \frac{\int_{80 \text{ms}}^{\infty} p_f^2(t) \, dt}{\int_{5 \text{ms}}^{80 \text{ms}} p_o^2(t) \, dt} \right] \text{[dB]},$$

where $p_{10}(t)$ is the IR of the sound source normalized at a distance of 10 m away in a free field.

### 2.3 Measurement Uncertainty

A number of studies have shown that there is a high degree of measurement uncertainty in room acoustics metrics obtained from room IRs [15-25]. Specific sources of uncertainty and studies between measurement teams are summarized below. A common method to evaluate uncertainty is to compare measurements in terms of just noticeable differences (JNDs). The JND for each room acoustics parameter is included in the Annex A of ISO 3382 [43]: 5% for T30 and EDT, 1 dB for C80, 1 dB for G, 0.05 for $J_L$, 0.05 for definition (D), and 10 ms for center time ($T_s$); the JND for $L_i$ is not known. For the purposes of this study, the JND for G will be used for $L_i$.

The contributions of different sources of uncertainty to the overall measurement uncertainty has been studied in Ref. [47]. The main contributions to measurement uncertainty are source position and orientation, microphone placement and orientation, source directivity, microphone directivity, and measurement hardware frequency response. Source directivity, in particular, has been shown to be a significant portion of the measurement uncertainty as a result of non-uniform source directivity. The most common sound sources used in room acoustics are dodecahedron loudspeakers, which typically become directional above approximately 1 kHz. Therefore, the orientation of the source can yield different results in room acoustic metrics [48, 49]. A second major contributor is microphone placement, where measures can vary widely even within a single seat location [50, 51]. Additional sources of uncertainty include ambient room conditions (i.e. temperature and humidity), evaluation methods (e.g. different signal processing and filtering methods), room noise, and equipment noise.
Studies comparing metrics calculated from IRs obtained from different measurement teams show differences that exceed the JND of each metric in most cases [20-22]. One of the earliest studies comparing the results from four measurement teams showed that the standard deviation across the teams were around 5% to 10% for T30, EDT, D, and T5, and around 0.5 dB for C80 and G from 1 kHz to 4 kHz, which are all on the order of 1-2 JNDs [52]. The largest differences tended to occur in the 125 Hz octave band. Additionally, larger differences were found in LF measurements with differences up to 4 JNDs at 1 kHz.

The first phase of a second study – the third round robin on room acoustics simulation programs – was to collect measurement data on the space that was to be modeled [53]. T30, EDT, C80, and G measurements all showed differences well above 1 JND with the largest differences in the 125 Hz octave band. Again, the largest differences were found in the parameter Lf, which were on the order of 3 to 5 JNDs in various octave bands and receiver positions. As part of the third round robin study, some follow-up measurements using three figure-8 microphones of the same make and model (Neumann KM86) revealed significant differences in measurements taken with the microphones at different orientations (i.e. rotated 180 degrees). One possible source of this measurement error was hypothesized to be due to changes in the microphone sensitivity of each diaphragm due to aging.

A third study compared measurements made in opera houses using different measurement hardware and excitation techniques [54]. Differences were found using different types of sources, e.g. different types of dodecahedron loudspeakers and impulsive sources such as balloons or starter pistols. Measurements were compared using an omnidirectional microphone and two sets of binaural microphones in the same dummy head. The two binaural microphones had differences in C80 up to 0.5 dB at low frequencies, which is below 1 JND. Measurements made comparing the binaural microphones to the omnidirectional microphone showed differences greater than 1 dB for C80 (1 JND) and 15% for reverberation time (3 JNDs).

A handful of studies have been conducted in which the same measurement team evaluated measurements using different microphones. In these studies, measurement uncertainty caused by the source was significantly reduced because researchers used the same source for all of the measurements. Additionally, uncertainty due to microphone placement should have also been reduced since placement within a team is likely to be more consistent than across teams.
Therefore, in these studies a majority of the uncertainty should be due to the differences in microphone response. A study in 2006 compared parameter differences for measurements made with four different types of microphones [55]. Standard deviations for the 10 measurements made with each microphone type were small, indicating that measurements were consistent and repeatable, but differences between microphone types were found to be approximately 1 JND for T20 and G, and between 1 and 2 JNDs for LF. A second study compared the results of spatial measurements made from a figure-8 and omnidirectional microphone pair and an intensity probe made with two omnidirectional capsules [56]. Differences in JLF between the two microphone types were between 1 to 4 JNDs in various octave bands. The authors of the study speculated that most of the differences in the measurements were due the directivity of the microphones, which were found to deviate from the ideal patterns. A more recent case study was conducted to further evaluate the measurement uncertainty of the spatial measures LJ and JLF in terms of microphone orientation, spacing between the microphone pair, and microphone type [57]. A total of five different makes and models of figure-8 microphones were evaluated by taking measurements in a small lecture hall with about 100 seats. The average differences due to microphone spacing, which varied between 64 to 152 mm, and microphone orientation, were found to be relatively small for LJ and JLF, with differences below 1 JND, 0.2 dB and 0.02, respectively. On the other hand, the effect of microphone type was more significant for LJ, with variations on the order of 1.5 dB, but approximately 1 JND for JLF.

Based on findings from the aforementioned studies, the goal of ensuring differences in microphone types that are under 1 JND may be difficult to achieve in practice. In most of the studies involving measurement uncertainty, measurements made in the same location with different microphones or between different microphone teams exceeded 1 JND, which would indicate a perceptible difference between the measurements even though there may be none. In general, findings show that there is more uncertainty associated with spatial measures using a figure-8 microphone than in metrics only involving omnidirectional microphone measurements [20-25].

2.4 SPHERICAL MICROPHONE ARRAY PROCESSING AND BEAMFORMING

As shown in [26], the pressure on the surface of a sphere due to an incident plane can be represented as an infinite sum of spherical harmonics:
\[ p(r, \theta, \phi, t) = P_0 4\pi \sum_{n=0}^{\infty} i^n b_n(ka) \sum_{m=-n}^{n} Y_n^m(\theta, \phi)Y_n^m*(\theta_i, \phi_i) e^{i\omega t}, \quad (2-3) \]

where \( p \) is the total sound pressure, \( P_0 \) is the pressure amplitude, \( i \) is the imaginary number \( \sqrt{-1} \), \( \theta \) is the elevation angle, \( \phi \) is the azimuthal angle, and the direction of the incident wave is \((\theta_i, \phi_i)\), and * denotes the complex conjugate. \( Y_n^m \) are the spherical harmonics of order \( n \) and degree \( m \), which are defined as:

\[
Y_n^m(\theta, \phi) = \sqrt{\frac{(2n+1)(n-m)!}{4\pi(n+m)!}} p_n^m(\cos \theta) e^{im\phi}. \quad (2-4)
\]

The coefficients \( b_n \) are known as the plane wave modal coefficients and are dependent on the array geometry and the boundary conditions at the surface of the sphere. For a rigid sphere, the coefficients are [38]:

\[
b_n(ka) = j_n(ka) - \frac{j_n'(ka)}{h_n^{(2)'}(ka)} h_n^{(2)}(ka), \quad (2-5)
\]

where \( j_n \) are spherical Bessel functions of order \( n \), \( h_n^{(2)} \) are spherical Hankel functions of the second kind of order \( n \), and \( ' \) signifies a derivative with respect to the argument.

The spherical harmonics form an orthonormal basis set, meeting the orthogonality condition:

\[
\int Y_{n'}^{m'}(\theta, \phi) Y_{n}^{m*}(\theta, \phi) = \delta_{n'n} \delta_{mm'}, \quad (2-6)
\]

where \( \delta \) is the Kronecker delta function. Because of the orthogonality property of the spherical harmonics, the spatial Fourier coefficients for the spherical harmonics \( \tilde{P}_{nm}(ka) \) can be obtained by applying weights to each microphone signal and summing the signals together:

\[
\tilde{P}_{nm}(ka) = \frac{1}{b_n(ka)} \sum_{s=1}^{S} P_s(ka)Y_{n}^{m*}(\theta_s, \phi_s), \quad (2-7)
\]

where \( P_s \) is the complex pressure in the frequency domain measured at microphone \( s \), obtained by taking a discrete Fourier transform (DFT) of each microphone signal, and \((\theta_s, \phi_s)\) is the location of the microphone on the sphere.
The spatial Fourier components can be weighted and combined to beamform different directional patterns in the spherical harmonics domain. A plane wave decomposition can be performed by summing the components together and setting the beamforming weights to unity [58]:

\[ P(\theta_l, \phi_l) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \tilde{P}_{nm}(ka) Y_n^m(\theta_l, \phi_l), \]  

where \((\theta_l, \phi_l)\) is the look direction of the beam (the direction in which the main lobe of the beam pattern is oriented). To calculate the parameters in ISO 3382 [43], the omnidirectional (monopole) and figure-8 (dipole) components must be extracted from spatial Fourier coefficients. The omnidirectional component can be extracted using the zeroth order spherical harmonic component:

\[ P_o = Y_0^0 P_{00} = \frac{1}{(4\pi)^2 b_0(ka)} \sum_{s=1}^{S} P_s(ka). \]  

The figure-8 component can be extracted by weighting and summing the first order spherical harmonic components:

\[ P_f = \sum_{m=-1}^{1} P_{1m} Y_1^m(\theta_l, \phi_l) \]
\[ = \frac{1}{b_1(ka)} \sum_{m=-1}^{1} Y_1^m(\theta_l, \phi_l) \sum_{s=1}^{S} P_s(ka) Y_1^m(\theta_s, \phi_s), \]  

where \((\theta_l, \phi_l)\) is the look direction of the beam (where the dipole is a maximum). The dipole pattern must be steered in the proper direction to calculate room acoustics parameters. The look direction should be chosen such that the elevation angle is pointed toward the source, and the azimuthal angle is the direction of the source plus 90 degrees. This process will align the null plane of the dipole pattern with the sound source. The time domain omnidirectional and figure-8 signals are calculated using an inverse DFT of \(P_o\) and \(P_f\), respectively. Room acoustics metrics can then be calculated using the beamformed IRs.
The theoretical weights used in the beamforming analysis, given in Eqns. (2-7) through (2-10), assume that the spherical microphone array used in this study is close to ideal, i.e., the geometry is a uniformly sampled rigid sphere and the array elements are perfectly matched in magnitude and phase. The theoretical weights were found to work well in this application and the desired directivity patterns were achieved (see Section 2.5.2). However, any phase mismatches or calibration differences can cause a degradation in the achieved beampattern. If mismatches between microphone capsules are present, then it may be advantageous to use weights based on measurements of the array rather than theoretical weights [59] [60]. Applying this technique can improve the performance of the beamformer in order to achieve a more ideal directivity pattern.

2.5 MEASUREMENTS

2.5.1 Measurement Hardware

The spherical microphone array used in this study was an Eigenmike em32 array made by mh Acoustics, as shown in Figure 2-1a [61]. This array consists of 32 omnidirectional electret microphone capsules distributed uniformly on an 8.4-cm diameter rigid sphere. The microphone capsules are individually calibrated to account for magnitude differences between capsules. The upper frequency limit of the spherical array due to spatial aliasing is approximately 8 kHz. The Eigenmike system includes the Eigenmike Interface Box (EMIB), which is used to convert the Eigenmike data into a standard Firewire interface that can be controlled via an ASIO driver. The EMIB also contains a digital output, which allows the EMIB to send the excitation signals to the sound source. The ASIO driver and digital output make it possible to take IR measurements with synchronized inputs and outputs using commercially available room acoustics measurement software. This system allows for simultaneous measurements of the 32 individual microphone channels of the Eigenmike.

Traditional IR measurements were made using a Brüel & Kjær (B&K) Type 4192 12.7 mm (0.5 inch) omnidirectional microphone and a Sennheiser MKH 30 figure-8 microphone. These microphones were chosen because they are representative of a typical measurement microphone and studio-grade figure-8 microphone, respectively. The omnidirectional and figure-8 microphones were placed on the same microphone stand 7.6 cm apart from each other and measured simultaneously, as shown in Figure 2-1b. This spacing was chosen to allow for the
microphones to be adequately far enough apart so as to minimize the effects on the other microphone, but close enough so that they were measuring approximately the same point in space [57].

![Microphones](image)

*Figure 2-1: Microphones used in this study. (a) Eigenmike em32 spherical microphone array and (b) a Brüel & Kjær (B&K) Type 4192 omnidirectional and Sennheiser MKH30 figure-8 microphone pair.*

A B&K Type 4292-L OmniPower Sound Source dodecahedron loudspeaker was used to excite the room for the IR measurements, which was placed at the center of the stage, and driven with a Crown K2 power amplifier. For the Eigenmike configuration, the Eigenmike Interface Box (EMIB) was used as the audio interface. The digital output of the EMIB was sent to an RME Babyface, which was used as a digital-to-analog converter to send the excitation signals to the amplifier. For the traditional IR measurements, the RME Babyface audio interface was used to send excitation signals to the amplifier and to receive microphone signals.

In order to place the microphones quickly and reliably, a custom microphone stand was built that could be used for both microphone configurations: the Eigenmike array, and the omnidirectional and figure-8 pair. As shown in Figure 2-2 (and in a video posted online [62]), the microphone stand consists of a base that sits on the floor in front of the seat with adjustment screws that allow the stand to be leveled, a body that raises and lowers to adjust the height, and an arm which protrudes over the seat and places the microphone in the position of a listener’s head. The stand features many adjustment points that allow for precise and accurate positioning in the x, y, and z individually without altering the other dimensions. After
microphone placement, the microphone can be rotated to align the microphone with the source while maintaining the position.

![Custom microphone stand used for accurate and precise placement of microphones. The photo on the left shows the stand being used to place the microphone and the one on the right shows the microphone in the final position.]

**Figure 2-2:** Custom microphone stand used for accurate and precise placement of microphones. The photo on the left shows the stand being used to place the microphone and the one on the right shows the microphone in the final position.

2.5.2 Anechoic Chamber Directivity and Frequency Response Measurements

Dodecahedron loudspeakers that are commonly used for room acoustics measurements are typically about 40 – 50 cm in diameter, which limits the omnidirectional radiation up to approximately 1 kHz. Consequently, ISO 3382:2009 requires the measurement of the sound source as a function of azimuthal angle in a free field to calculate certain metrics from the IR, specifically $G$ and $L_I$ [43]. IR measurements were made in an anechoic chamber at 12.5 degree increments around the B&K loudspeaker at a distance of 3.55 m from the source using a B&K 4191 12.7 mm (0.5 inch) free-field microphone as specified in the standard to account for the variation in directivity of the source as a function of azimuthal angle. The 29 measurements were energy-averaged in the frequency domain and normalized to a distance of 10 m from the source. The averaged result was used as the normalization IR for $G$ and $L_I$. The averaged result was also inverted and used to generate a filter to equalize the frequency response of the loudspeaker.
Anechoic measurements were also conducted to evaluate the sensitivity, frequency response, and directivity of both the Eigenmike array and the Sennheiser MKH30. The sensitivity was measured by playing a 1 kHz tone over the dodecahedron loudspeaker and measuring it with both microphones and a calibrated sound analyzer, B&K type 2250. The directivities were evaluated by measuring the IR of the microphones as a function of angle with a stationary loudspeaker placed 2 m from the microphone. The IRs of the microphones were measured every 3 degrees. The Sennheiser MKH30 and spherical array beamformed dipoles both have the correct directivity pattern with two clear nulls and a matching response at both maxima, as shown in Figure 2-3.

![Figure 2-3: Microphone directivity plots for the Sennheiser MKH30 figure-8 microphone (left) and Eigenmike em32 beamformed dipole pattern (right).](image)

The deviation of the microphone directivity patterns from the ideal patterns for the Sennheiser MKH30 and Eigenmike em32 microphones are shown in Figure 2-4a and 4b, respectively. For the two figure-8 patterns, the deviations are a fraction of 1 dB, with the exception of angles close to the nulls, where the majority of the sound is being rejected but the rejection is not infinite. Additionally, the directivity pattern for the Eigenmike em 32 beamformed omnidirectional pattern is very close to ideal (Figure 2-4c) with average deviations also on the order of a fraction of 1 dB.
Figure 2-4: Deviations from ideal polar patterns: Sennheiser MKH30 figure-8 microphone directivity versus a perfect dipole (a), Eigenmike em 32 beamformed dipole versus a perfect dipole (b) and Eigenmike em32 beamformed omnidirectional pattern versus a perfect omnidirectional (c) pattern.

2.5.3 Impulse Response Measurements in Eisenhower Auditorium

IR measurements were taken in the Eisenhower Auditorium (2500 seats) located on The Pennsylvania State University campus in University Park, PA, USA. The auditorium is a multipurpose venue used for a wide range of performance types and lectures. The measurements were taken with a moveable orchestra shell and overhead reflectors in place. The dodecahedron omnidirectional source was placed in the center of the stage for all measurements. Six receiver locations were chosen in the hall as shown in Figure 2-5: two on the main floor (R1 and R2), two on the grand tier level (R3 and R4), and two on the balcony level (R5 and R6).
At each receiver position, the IR was obtained using the 2-channel FFT correlation measurement technique as implemented in the software program EASERA [63]. The excitation signal was a sine sweep with a pink-weighted spectrum. Each measurement was taken using 10 sweep averages, along with an additional pre-sweep to avoid audio artifacts. Measurements were taken using both microphone setups: the Eigenmike array, and the omnidirectional and figure-8 microphone pair. The center of each microphone setup (either the center of the Eigenmike array or the center of the two discrete microphones) was placed 20 cm from the seat back, 70 cm above the seat bottom, and halfway across the width of the chair, which were either 46 or 48 cm depending on the location in the hall. The Eigenmike array was aligned so that the 0 degrees azimuth direction was aligned with the source. The omnidirectional and figure-8 pair was aligned such that the null plane of the figure-8 microphone was oriented vertically toward the loudspeaker. The microphone alignments were done by eye. To evaluate measurement repeatability, a total of three measurement sets were taken for each microphone setup at each receiver location. In between sets, the custom microphone stand was removed and the various adjustment points were loosened and randomly repositioned. The stand was then replaced in the receiver location and re-aligned toward the source to re-measure the IRs for each microphone.
2.6 DATA PROCESSING

2.6.1 Microphone frequency response compensation

The frequency response of the omnidirectional sound source, the Sennheiser MKH30 microphone, and the Eigenmike array capsules were not found to have flat frequency responses, and thus required the application of equalization filters.

The Eigenmike random incidence frequency response was measured by playing uncorrelated pink noise over a nearly spherical loudspeaker array consisting of 30 loudspeakers in an anechoic chamber with the Eigenmike placed in the center of the array [36]. The monopole component of the Eigenmike measurement was extracted using Eqn. (2-9) and was then third-octave smoothed in the frequency domain. The same noise signal was measured using a B&K 4942 diffuse-field 12.7 mm (0.5 inch) omnidirectional microphone placed in the center of the array to use as a reference, and then this signal was also one-third-octave smoothed in the frequency domain. The desired response for the Eigenmike equalization filter was generated by: (1) taking the smoothed spectrum of the B&K microphone measurement and dividing by the smoothed spectrum of the Eigenmike monopole measurement, and (2) creating a 4096 point linear-phase FIR filter from the desired response using the MATLAB function “fir2” [64]. The filter was normalized so that the equalization applied unity gain at 1 kHz. This normalization convention allows for the equalization filter to be used after a single-tone 1 kHz calibration. The magnitude response of the Eigenmike filter along with the target filter response are shown in Figure 2-6 as the solid and dashed blue lines, respectively.

Although the frequency response of the Sennheiser MKH30 is reasonably flat at mid-band frequencies, comparing the response of the MKH30 to the response of the B&K measurement microphone or the Eigenmike revealed that the response deviates by ±2 dB at high frequencies, and the response begins to gently roll off at low frequencies below 200 Hz. An equalization filter was generated for the MKH30 using the anechoic chamber measurements described in Section 2.5.2. The desired filter response was generated by taking the one-third octave band smoothed spectrum of the free-field measurement of the MKH30 directly on-axis, and dividing by the third-octave smoothed spectrum of the Eigenmike beamformed dipole on-axis with the Eigenmike’s equalization filter applied as shown in Figure 2-6 as the dashed red line. A 4096 point linear-phase filter was generated using the same MATLAB function “fir2” as before, shown
in Figure 2-6 as the solid red line. The on-axis free field measurements were used to equalize this microphone instead of diffuse field measurements.

![Graph showing Eigenmike and Sennheiser equalization filters]

*Figure 2-6: Eigenmike (blue) and Sennheiser MKH30 (red) equalization filter magnitude response. Target responses are shown as dashed lines, and realized filters fit to the target responses are shown as solid lines. The Eigenmike target and actual filter are nearly identical, which is why the dashed line is difficult to see in the figure.*

### 2.6.2 Rotating the dipole

For the figure-8 IR, the null plane should be pointed directly at the source. Alignments of the microphones were done by eye for this study. Alignment for the receiver locations toward the back of the hall and on upper seating levels can be especially difficult due to the source-receiver distances being the order of 35 m. To correct for any rotational misalignments in the spherical microphone array measurements, the direct sound can be beamformed to determine the actual source direction (or the desired look direction), and the dipole beam can be steered to properly align the null plane.

For each Eigenmike IR, a rectangular window was applied in the time domain to extract only the direct sound from the IR, and a third-order plane wave decomposition was conducted using Eqn. (2-8) for azimuthal look directions from -90° to 90° at 2 kHz. The 2 kHz frequency bin was chosen
because the inverse of the plane wave modal coefficients (Eqn. (2-5)) are relatively small for the third-order coefficient at that frequency, and the high orders do not require a large gain. The angle with the maximum energy was found and that angle was used to rotate the dipole beam in azimuth.

2.6.3 Room acoustic metric calculation

In order to compare the measured IRs from the spherical array and the omnidirectional and figure-8 pair, the following metrics were evaluated: reverberation time (T30), early decay time (EDT), clarity index (C80), strength (G), early lateral energy fraction (J_{LF}) and late lateral energy level (L_l). For the measures using the spherical array, the omnidirectional (monopole) and figure-8 (dipole) IRs were extracted from the raw IRs using Eqns. (2-9) and (2-10), respectively.

The omnidirectional parameters (T30, EDT, C80, and G) were calculated from the IRs measured from each microphone configuration, i.e. the monopole component of the Eigenmike, and the omnidirectional microphone. J_{LF} was calculated using the monopole and dipole components of the Eigenmike measurements, and the omnidirectional and figure-8 microphone pair. L_l was calculated using the dipole component of the Eigenmike measurements and the figure-8 microphone, respectively, with both configurations using the same free-field measurement of the sound source described in Section 2.5.2.

The measured IRs were analyzed by calculating the metrics following ISO 3382 [43] using MATLAB [64]. First, each IR was filtered to equalize the frequency response of the microphones and measurement loudspeaker (see Section 2.6.1). It is important to note that for the energy parameters, time-windowing was performed before octave-band filtering as defined in the standard. The order of this processing ensures that the group delay of the octave-band filters does not affect the energy integration limits. The results are presented as differences in the metrics, and compared with the just noticeable difference (JND) of the metric.

2.7 RESULTS

2.7.1 Measurement Repeatability of the Two Microphone Configurations

In this study, the largest source of measurement uncertainty is due to precise and accurate microphone placement in the exact same location for each measurement. In order to quantify the uncertainty due to microphone placement for each metric calculated for both microphone
configurations as described in Section 2.5.3, the standard deviation across the three repetitions for each measure at each receiver location was obtained. The standard deviations were then compared to the JND of the specific metric as listed in Section 2.2.

The standard deviations of the three repetitions for the six metrics measured with both microphone configurations for all receiver positions were calculated, and are shown in Figure 2-7. The thick-dashed red lines on each plot represent the respective 1 JND for each metric. In particular, the average standard deviations for each metric are as follows: 0.087 JNDs for T30, 0.19 JNDs for EDT, 0.085 JNDs for C80, 0.073 JNDs for G, 0.092 JNDs for JLF, and 0.069 JNDs for LJ. The differences were all found to be less than 1 JND, with a single exception of an EDT data point at 4 kHz at R3. Of the six metrics evaluated in this study, EDT was found to have the highest standard deviations relative to 1 JND. This result is consistent with a prior study [53] and appears to indicate that EDT is the most susceptible to small differences in microphone placement since this metric is dependent on the early sound field. The largest standard deviations for EDT were found in the 2 kHz and 4 kHz octave bands. This finding makes sense because at high frequencies the small microphone misalignments can be large relative to a wavelength.

Since prior research has indicated that there is a large degree of uncertainty with spatial measures [21,23-25], it was hypothesized that the same would be true of the data in this study. However, as shown in Figure 2-7, the spatial measures were found to be as consistent as T30, C80 and G.
Figure 2-7: Standard deviation of the three repeated measurements for each metric at each receiver location for the omnidirectional and figure-8 microphone pair (solid lines) and the Eigenmike array (thin-dashed lines) configurations. The thick-dashed red lines on each plot represent the respective 1 JND for each metric. All standard deviations were found to be well below 1 JND for each metric and all receiver positions, with the exception of a few cases for EDT.
The measurement repeatability of the two microphone configurations were compared using the standard deviations of each calculated metric over the three measurements (solid versus thin-dashed lines in Figure 2-6). For T30, G, J_{LE}, and L_{J}, the standard deviations were found to be similar with both microphone configurations. However, at high frequencies the two metrics that are the most susceptible to uncertainty due to microphone placement misalignments, EDT and C80, had larger standard deviations in the Eigenmike measurements. It was, in general, more difficult to align the Eigenmike in the physical space than traditional the traditional microphone pair, since the array is quite a bit larger than traditional measurement microphones and the round geometry makes it difficult to align the microphone’s center or edges. The larger standard deviations confirm that the spherical array was more prone to slight alignment errors than the traditional microphones.

For the Eigenmike measurements, the dipole pattern was rotated so that the null plane was perfectly aligned with the direct sound. The authors hypothesized that the rotational alignment of the figure-8 microphone or the dipole pattern would be a significant source of uncertainty, which would result in a lower standard deviation for the spatial parameters L_{J} and J_{LE} measured with the Eigenmike than measured with the Sennheiser microphone. However, the standard deviations measured with the Eigenmike and the Sennheiser microphone do not appear to be appreciably different for the spatial parameters. Therefore, most of the variation is likely due to microphone placement and not due to rotational alignment. To further investigate the effect of a rotational misalignment, the Eigenmike IRs at R3 were beamformed with the null plane oriented in azimuth angles from -5 to +5 degrees in 0.1 degree steps, where the direct sound was incident at 0 degrees. The 5 degree misalignment caused an error in J_{LE} of up to 0.02, and an error in L_{J} of up to 0.15 dB, which are both within 1 JND.

In addition to evaluating repeatability using standard deviations, a test-retest reliability analysis [65] was performed using the statistical software SAS [66]. For each microphone configuration, metric, and octave band, a correlation analysis was run between the three measurement repetitions over the six receiver locations. The measure of test-retest reliability is Cronbach’s alpha coefficient, where the coefficient is a value between 0 and 1 and a higher value indicates more reliability. In general, a value of 0.7 or 0.8 indicates a high degree of reliability [65]. The majority of the 72 output coefficients were above 0.99. The two data points that are below 0.99
are: for EDT at 4 kHz measured with the Eigenmike (0.98), and T30 at 4 kHz measured with the B&K microphone (0.93). The results of this statistical test provide more evidence that the measurements are repeatable.

In summary, the overall measurement uncertainty was found to be less than 1 JND for all six metrics for both microphone configurations. Given the low amount of uncertainty and the consistent results across both microphone configurations, the following analysis, described in the next section, was carried out to compare the calculated metrics based on the IRs of the two microphone types. This analysis was conducted to determine if a spherical array can be used instead of the traditional microphone pair to measure typical room acoustics metrics and is described in the next section.

2.7.2 Differences in Measured Room Acoustics Metrics between Microphone Configurations

To compare the measurements obtained using both microphone configurations as described in Section 2.5.3, the calculated metrics for each receiver position were averaged over the three repetitions. These averages were then directly compared between the two configurations by subtracting the metrics obtained using the spherical microphone configuration from the measurements taken using the traditional microphone configuration. These differences were compared to the JNDs for each metric.

A plot of the differences for each parameter is shown in Figure 2-8. The differences in T30 and EDT are shown as percentages and the remaining metrics are shown as absolute differences in order to be able to compare measured differences to their respective JNDs, indicated by a thick-dashed red line. The average differences for all of the measures are as follows: 0.18 JNDs for T30, 0.71 JNDs for EDT, 0.27 JNDs for C80, 0.25 JNDs for G, 0.28 JNDs for J_{LF}, and 0.26 JNDs for L_{J}.

The differences across the microphone configurations for all six metrics for all six receiver locations are within 1 JND with the exceptions of R3 at 2 kHz for EDT and C80, and a few other data points for EDT, which are discussed below.

Of particular interest for this study were the differences in the spatial measures, J_{LF} and L_{J}, obtained from the two microphone configurations considering previous studies showed a large amount of uncertainty across different microphones and different measurement teams [53, 55, 56, 57]. The results of this study show that in all cases the differences between the spherical
array measurements and traditional measurements are within 1 JND for the spatial measures. These results indicate that the metrics obtained from the IRs of either microphone configuration are perceptually identical, and that the spherical microphone array will yield similar results to measurements made with a traditional microphone configuration.

To reduce the differences in the spatial measures $J_{1/3}$ and $L_n$, the beamformed Eigenmike dipole was rotated in an attempt to match any rotational misalignments that may have been present in the Sennheiser figure-8 measurements. The agreement between the two microphone configurations did not significantly improve after rotating the Eigenmike dipole in azimuth angles from -5 to +5 degrees in 0.1 degree steps. In fact, the variation decreased in some octave bands, but increased in other octave bands, which indicates that the rotational misalignments in these measurements are not the dominant source of the differences between microphone configurations.
Figure 2-8: Differences between the omnidirectional and figure-8 microphone pair and the Eigenmike array configurations for all six metrics measured at all six receiver locations. The thick-dashed red lines on each plot represent the respective 1 JND for each metric. All differences were found to be within 1 JND, with the exceptions of a single point in C80, and several points in EDT at high frequencies.
Although the differences between the two microphone configurations for most metrics are within 1 JND, 8 of the 216 data points (3.7%) do exceed this threshold. Most of the differences were only slightly more than the 1 JND threshold. In the 2 kHz octave band, these differences were found in C_80 at R3 and EDT and R6. In the 4 kHz octave band, EDT differences slightly more than 1 JND were found at R2, R4, R5, and R6.

The single data point with a large discrepancy is the 4 JND difference in EDT in the 2 kHz octave band at R3. A more detailed analysis of the measured IRs from the two microphone configurations at R3 was conducted to better understand this result. As shown in Figure 2-9, the differences in EDT can be seen by comparing the energy decay curves of the two microphone configurations. The decay curves are both normalized to a maximum level of 0 dB. The slopes of the first 10 dB of the decay curves are very different primarily due to two spikes in the curves, as highlighted with ovals in Figure 2-9. These spikes may be a result of the physical location of R3 in the side portion of the first balcony (see Figure 2-5). This location is especially close to reflecting surfaces, including the side wall and the balcony ceiling. Since EDT is strongly influenced by early reflections, this measure may be particularly sensitive to the precise microphone position at this receiver location.
Figure 2-9: The energy decay curves and associated early decay time slope fits for R3 in the 2 kHz octave band shown for the omnidirectional B&K microphone (blue) and Eigenmike array (red). Note the spikes in the curve as denoted with the orange ovals, which are likely due to differences in where the centers of each microphone were positioned.

One possible reason for the differences in EDT, and to a lesser extent C80, is that the center of (in other words the space between) the B&K and Sennheiser pair was in the same position as the center of the Eigenmike array. As a result, the physical location of the center of the B&K microphone was 3.8 cm in front of the center of the Eigenmike array. Although this location is still inside the 8.4-cm sphere that the Eigenmike array physically occupies, the measurement of EDT could vary significantly over the sphere of the array. To investigate this possibility, the EDT was calculated for each of the 32 individual omnidirectional Eigenmike capsules separately in the 125 to 4,000 Hz octave bands at R3, as shown in Figure 2-10. The variation in EDT across the 32 capsules exceeded the 5% JND in all octave bands. It was expected that the EDT values would have large variations across the 32 individual microphones at high frequencies due to the shadowing effect of the rigid sphere. However, in the 125 Hz and 250 Hz octave bands where the shadowing effect is negligible (i.e. very little energy is scattered from the 8.4-cm sphere), the range in measured EDT values still exceeds the 5% JND. This finding provides more evidence that the EDT measurement is very sensitive to the physical microphone position. Given that the EDT
values in all six octave bands measured with the B&K microphone all fall within the range of EDT values measured with the individual Eigenmike capsules, the differences in measured values between the Eigenmike and B&K microphone are most likely caused by the offset of 3.8 cm in the position of the center of the microphones.

Figure 2-10: Early decay time calculated for Eigenmike individual microphone capsules (green), and early decay time calculated for the omnidirectional B&K microphone (blue) at R3.

2.8 CONCLUSIONS

The purpose of this study was to compare the room acoustics metrics defined in ISO 3382 [43] when obtained from IRs measured with a conventional microphone configuration, an omnidirectional and figure-8 pair, to those measured with a spherical microphone array. IR measurements were obtained in a 2500-seat auditorium using a B&K 4192 12.7 mm (0.5 inch) microphone and a Sennheiser MKH30 microphone, and an Eigenmike em32 spherical microphone array. Six receiver locations were measured throughout the hall, with three repetitions at each receiver to evaluate the uncertainty in the positioning of the microphones at each location. A custom microphone stand was utilized to ensure accurate and precise placement for each repetition. The room IRs were filtered to equalize the frequency response of the microphones and sound source, and metrics were extracted from the IRs using MATLAB. The
measures evaluated as part of this study were reverberation time (T30), early decay time (EDT), clarity index (C80), strength (G), early lateral energy fraction (J\text{LF}), and late lateral energy level (L\text{L}).

In order to evaluate measurement repeatability, the standard deviations of the three repetitions were calculated for each metric at each receiver location for both microphone configurations. The standard deviations were evaluated across the metrics and microphone configurations, and compared to the JND of each measure as defined in the Annex of ISO 3382 [43]. On average, across all metrics in the octave bands from 125 to 4000 Hz at all six receiver locations, the average standard deviations were 0.099 JNDs. One exception was found for a single standard deviation of 1.2 JNDs for EDT in the 4 kHz octave band at R3 measured with the Eigenmike. The deviation likely occurred due to slight misalignments in re-positioning the spherical array during repetitions. A test-retest reliability analysis was performed, and Cronbach’s alpha coefficient was found to be above 0.99 for the majority of the measurements (above 0.8 is considered to be the criteria for reliability), confirming that the three repetitions were consistent. Overall, microphone placement was found to be repeatable between the three repetitions of each measurement, which indicates that the measurement uncertainty is relatively low, and it is valid to compare the measurements made with each microphone configuration.

Differences in the four omnidirectional and two spatial metrics evaluated in this study, as obtained from the IRs between the two microphone configurations, were calculated at each receiver location. The differences in the measures were smaller than 1 JND for the majority of the metrics, with the exception of C80 at one single receiver position and octave band, and EDT in several receiver positions at high frequencies (2 kHz and above). These metrics were found to be very sensitive to small changes in position of the microphones since they are dependent on the early sound field. The measurement differences in EDT and C80 are likely due to the physical center of the B&K omnidirectional microphone being 3.8 cm in front of the physical center of the Eigenmike array, which was done in order to keep the center of the microphone pair and the center of the Eigenmike array coincident. Based on previous research, the authors’ hypothesis was that the largest variation between microphone configurations would be found for the spatial metrics L\text{L} and J\text{LF}, but instead these metrics were found to have smaller differences than EDT and C80 relative to 1 JND.
Overall, the agreement in the calculated metrics obtained from the IRs of the two microphone configurations is much better than the agreement reported in previous studies for differing microphone types [55, 56, 57]. The measurement consistency in the results between the two microphone types was achieved because of matching directivity patterns between the microphones and beamformed signals, a flat frequency response for each microphone achieved by filtering to equalize the microphone signals, and accurate and precise placement of the microphones in the seat locations. Further studies are encouraged that include a larger number of microphones. Based on the results of this case study, it is acceptable to measure room acoustic metrics with a spherical microphone in place of the traditional configuration of an omnidirectional and figure-8 microphone pair. This finding will enable the use of spherical microphone arrays to measure both existing room acoustics metrics and to develop new metrics using higher order beamformed directivities.

2.9 ACKNOWLEDGEMENTS

The authors wish to express their thanks to Mr. Tom Hesketh, for allowing access to Eisenhower Auditorium for taking the measurements. The authors also wish to acknowledge Matthew Neal, Matthew Kamrath, Martin Lawless, and Acadia Kocher for their assistance with the measurements; and Matthew Neal for demonstrating the procedure used to position the microphone using the custom microphone stand in the online video [62]. The authors would like to thank Bose Corporation in Framingham, MA, USA for use of their anechoic chamber to measure the directivities of the Sennheiser MKH30 microphone and Eigenmike array. This work was supported by National Science Foundation (NSF) award #1302741.
Chapter 3: An Investigation of Listener Envelopment Utilizing a Spherical Microphone Array and Third-Order Ambisonics Reproduction

This chapter will discuss spatial room IR measurements obtained in a concert hall, two subjective studies using the room IRs, an evaluation of several metrics used to predict LEV, and trends found from LEV ratings compared with an objective analysis of the sound field. This text is formatted as a manuscript which will be submitted for review and publication in a peer-reviewed journal.
An Investigation of Listener Envelopment Utilizing a Spherical Microphone Array and Third-Order Ambisonics Reproduction

ABSTRACT

Listener envelopment (LEV), the sense of being surrounded by the sound field, is a perception that has been found to be related to the overall room impression of a concert hall. The purpose of this study was to investigate the relationship between the perception of LEV and the direction and arrival time of energy from spatial room impulse responses (IRs). IRs were obtained in a 2,000-seat concert hall in several receiver locations and hall absorption settings using a 32-channel spherical microphone array. The IRs were analyzed using a 3rd order plane wave decomposition (PWD). Additionally, the IRs were convolved with anechoic music and processed for 3rd order Ambisonic reproductions and presented to subjects over a 30-loudspeaker array. Instances were found in which the energy in the late sound field did not correlate with LEV ratings as well as energy in a 70 to 100 ms time window. Follow-up listening tests were conducted with hybrid IRs containing portions of an enveloping IR and an unenveloping IR with crossover times ranging from 40-140 ms. Additional hybrid IRs were studied wherein portions of the spatial IRs were collapsed into all frontal energy with crossover times ranging from 40-120 ms. The tests confirmed that much of the important LEV information exists in the early portion of these IRs.

3.1 INTRODUCTION

An important aspect of the overall room impression of a concert hall is the spatial impression of the hall, which includes listener envelopment (LEV), the sense of being fully immersed in the sound field. The purpose of this study was to investigate the relationship between the perception of LEV and the direction and arrival time of energy in performing arts spaces utilizing measurements obtained with a compact spherical microphone array. The impulse response (IR) measurements made using such an array can be used for both an objective analysis of the sound field in full 3D via beamforming techniques [25] and subjective listening tests using 3D reproductions of the sound fields over a loudspeaker array via Ambisonics [27] [28].

Research aimed at understanding the spatial perception in performing arts spaces initially focused on the directional dependence of early reflections. Originally, the sense of spaciousness
was thought to be primarily associated with reverberation, but in the late 1960’s, it was found that the spatial impression was heavily influenced by early reflections [3]. It was proposed that the spaciousness depended on the arrival direction of early reflections, and that stronger early lateral reflections were related to a quality referred to as “spatial responsiveness” [6]. Further work led to the development of the objective metric Early Lateral Energy Fraction \( J_{LF} \) (prior notation \( LF \)), the ratio between the early lateral energy and the total early energy in the first 80 ms, which was found to be correlated with the subjective level of spatial impression [7]. A second metric that has been found to correlate with spatial impression is the interaural cross correlation coefficient (IACC), which is obtained from the cross-correlation of the left and right ears of a binaural IR [8].

More recently, it has been proposed that the spatial impression of a hall contains two distinct perceptions: apparent source width (ASW), which is the sense of how wide or narrow the sound image appears to a listener, and LEV, the sense of being immersed in and surrounded by the sound field [9]. ASW has since been shown to be related to early lateral reflections, which can be predicted using IACC and \( J_{LF} \), while LEV has been shown to be related to late lateral energy [10].

Seminal work on LEV was conducted by Bradley and Soulodre in the early 1990’s [10] [67]. Using five loudspeakers distributed in the front half of the horizontal plane, sound fields were generated with a small number of early reflections that were kept constant, and varied certain aspects of the late sound field: the reverberation time (T30), the early-to-late sound energy ratio (C80), and the strength of the late sound field (\( G_{\text{late}} \)). The angular distribution of the late sound was also varied, which was accomplished by playing the late sound either out of a single frontal loudspeaker, three frontal loudspeakers spanning 70°, or five frontal loudspeakers spanning 180°. A subjective study showed that the parameters with the highest correlation to LEV were angular distribution and overall late level. These results were used to develop a metric to predict LEV called late lateral energy level, \( L_J \) (prior notation \( LG \)) [10]:

\[
L_J = 10 \log_{10} \left[ \frac{\int_{0}^{\infty} p_L^2(t) \, dt}{\int_{0}^{\infty} p_{10}^2(t) \, dt} \right] \text{[dB]}, \tag{3-1}
\]
where $p_L(t)$ is the room IR measured with a figure-of-eight microphone, and $p_{10}(t)$ is the IR of the sound source normalized at a distance of 10 meters away in a free field. While a strong correlation was found between this metric and LEV, it should be noted that this study used a small number of loudspeakers spanning a limited angular area and the simulated sound fields may not have been representative of a real concert hall. Additionally, the range of $L_I$ values of the stimuli was greater than 20 dB, which is a much larger range than would be found in actual spaces – Bradley reports a range of approximately -7 dB to +6 dB [68].

Although work by Bradley and Soulodre indicated that late lateral energy is the component of the sound field with the highest correlation with envelopment [10] [67], a number of studies have shown correlation between listener envelopment and non-lateral sound and/or early reflections. One subjective study showed that adding front early reflections from above (i.e. ceiling reflections), increases LEV [12]. A second subjective study found that increasing the energy behind a listener will increase LEV and that early reflections can also contribute to the perception of LEV [13]. A third subjective study was conducted using a similar loudspeaker arrangement to Bradley and Soulodre’s with the addition of a loudspeaker placed directly behind the listener and a loudspeaker overhead [14]. By varying the distribution of the late sound, findings showed that late sound both above and behind the listener significantly affect LEV [14], which directly contradicts a prior study’s finding that reflections above and behind do not significantly impact envelopment [15]. The effect of the early sound on LEV was also investigated in a study using binaural stimuli in which the left ear signal was fed into the right channel and vice versa in order to increase interaural cross correlation [16]. Findings from that study showed that cross-mixing the channels to increase the interaural cross correlation in only the early part of the IR decreased LEV, in some cases more so than cross-mixing only the reverberant tail.

Currently, $L_I$ is the most commonly used objective metric to predict LEV, and is included in Annex A of the room acoustics measurement standard ISO 3382 [17]. However, several other objective metrics have been proposed to predict LEV. Unlike ASW, it has been found that energy fractions and cross-correlation metrics of the late sound such as late lateral energy fraction (LLF) [18] and late interaural cross correlation coefficient (IACC$_{L3}$) [19] are, by themselves, poor predictors of LEV and do not significantly vary between halls or within halls, although they have
been used as a component of other metrics. \(L_i\), for example, can be calculated from LLF and \(G_{\text{Late}}\). Beranek has also proposed an empirical formula to predict LEV objectively that is based on \(\text{IACC}_{L,3}\), strength (G), and clarity index (C80) \([21]\). Other metrics that have been suggested include front/back energy ratio \([22]\) \([23]\) (FBR, Eqn. 3-A1), and spatially balanced center time (\(\text{SBT}_S\), Eqn. 3-A5), which is based on the center time of spatial IRs weighted by the arrival direction \([24]\) (see Appendix A for additional information about these metrics). However, little information exists on the performance of these metrics.

To summarize, many studies have found that the early sound field and non-lateral energy can have an influence on the perception of LEV, which is not accounted for in the metric \(L_i\). Although several other metrics have been proposed to predict LEV, they have not been widely adopted in the architectural acoustics community, and there are only a limited number of studies that evaluate the performance of each of these metrics. Additionally, some of these metrics were designed using simulated sound fields that were reproduced over a limited number of loudspeakers, and may lack realism when compared to actual concert halls.

The purpose of this study was to investigate the relationship between the perception of LEV and the direction and arrival time of energy using measured spatial IRs. In this study, spatial room IRs were obtained in a 2,000 seat dedicated concert hall with a volume of 24,000 m\(^3\) (850,000 ft\(^3\)) and a maximum mid-frequency reverberation time of 2.8 s (Peter Kiewit Concert Hall in Omaha, NE, USA). An 8.4-cm (3.3-in) diameter spherical microphone array containing 32 omnidirectional microphone capsules was used to capture the directional IRs. The measured sound fields were analyzed by beamforming directional IRs and plotting the spatial distribution of energy as a function of time in different octave bands. Additionally, the IRs were convolved with an orchestral anechoic music excerpt and reproduced over an array containing 30 loudspeakers located in an anechoic chamber to conduct two listening tests. The first listening test contained the original IRs, as measured. In order to study the time dependence of the energy in the IRs related to LEV, the second listening test contained IRs that were modified to a) include time portions from two different measured IRs, one perceived with low LEV and one with high LEV, and b) remove the spatial properties of different time portions of the IRs. The approach used in this study is an improvement over previous envelopment research due to the
resolution of the beamformed directional IRs, and due to the accurate reproduction of measured sound fields.

3.1.1 Spherical Array Beamforming

Measurements obtained with a spherical microphone array can be beamformed in the spherical harmonics domain to generate directional IRs. The sound pressure on the surface of a sphere due to an incident plane wave can be represented in the spherical harmonics domain as [26]

\[
p(a, \theta, \phi, t) = P_0 4\pi \sum_{n=0}^{\infty} i^n b_n(ka) \sum_{m=-n}^{n} Y_n^m(\theta, \phi) Y_n^{-m*}(\theta_i, \phi_i) e^{i\omega t},
\]

where \( p \) is the total sound pressure, \( a \) is the radius of the sphere, \( \theta \) is the elevation angle, \( \phi \) is the azimuthal angle, \( P_0 \) is the pressure amplitude, \( i \) is the imaginary number \( \sqrt{-1} \), and \( k \) is the wave number. The direction of the incident wave is \((\theta_i, \phi_i)\), and \( * \) denotes the complex conjugate. The spherical harmonics, \( Y_n^m \), of order \( n \) and degree \( m \) are shown in Figure 3-1 and defined as:

\[
Y_n^m(\theta, \phi) = \sqrt{\frac{(2n + 1)(n-m)!}{4\pi(n+m)!}} p_n^m(\cos \theta) e^{im\phi}.
\]

![Figure 3-1: Spherical harmonic functions of order n and degree m, up to n = 3. For convenience, the real-valued spherical harmonics are shown, where red indicates a positive value and blue indicates a negative value.](image)
In Eqn. (3-2), the coefficients \( b_n \) are radial functions, which are often referred to as plane wave modal coefficients and are dependent on the geometry of the array used. For a rigid sphere, the coefficients are [38]:

\[
b_n = j_n(ka) - \frac{j_n'(ka)}{h_n^{(2)}(ka)} h_n^{(2)}(ka),
\]

where \( j_n \) are spherical Bessel functions of order \( n \), \( h_n^{(2)} \) are spherical Hankel functions of the second kind of order \( n \), and \( ' \) indicates a derivative with respect to the argument.

The spherical harmonics form an orthonormal basis set that satisfy the orthogonality relation:

\[
\int_0^{2\pi} \int_0^\pi Y_{nm}'(\theta, \phi) Y_{nm}^*(\theta, \phi) \sin\theta \, d\theta \, d\phi = \delta_{mm'} \delta_{nn'},
\]

where \( \delta \) is the Kronecker delta. Taking advantage of the orthogonality, the spatial Fourier transform can be applied to the microphone signals to transform them into the spherical harmonics domain. The spatial Fourier coefficients for the spherical harmonics \( \tilde{P}_{nm}(ka) \) can be obtained by applying weights to each microphone signal and summing the signals together. For a nearly-uniformly sampled sphere with \( S \) microphones, the Fourier coefficients as a function of frequency \( (f = \frac{ck}{2\pi}) \), where \( c \) is the sound speed) becomes [25]:

\[
\tilde{P}_{nm}(f) = \frac{1}{b_n(f)} \frac{4\pi}{S} \sum_{s=1}^S P_s(f) Y_n^{m*}(\theta_s, \phi_s),
\]

where \( P_s \) is the complex pressure in the frequency domain measured at microphone \( s \), obtained by performing a discrete Fourier transform (DFT) on each microphone signal, and \( (\theta_s, \phi_s) \) is the location of the microphone on the sphere. The spatial Fourier components can be weighted and combined to beamform a directional pattern of order \( N \):

\[
P(\theta_l, \phi_l) = \sum_{n=0}^N c_n \sum_{m=-n}^n \tilde{P}_{nm}(ka) Y_n^m(\theta_l, \phi_l),
\]

where \( c_n \) are weights that determine the shape of the beampattern, and \( (\theta_l, \phi_l) \) is the look direction of the beam in which the main lobe of the beampattern is oriented. For the specific
case of plane wave decomposition (PWD), Eqn. (3-7) can be used with the $c_n$ weights set to unity [58]. Directional room IRs can be obtained by beamforming the IRs measured with a spherical array.

### 3.1.2 Ambisonics Reproduction

Third order Ambisonics (more generally referred to here as Ambisonics) was utilized in this study to generate spatial reproductions of the measured data. Ambisonics is a spatial audio playback system originally developed by Gerzon in the 1970’s as a method to reproduce sound fields represented in the spherical harmonics domain [27]. Ambisonics initially only used the zeroth- and first-order spherical harmonic components and has since been extended to higher orders [28]. Ambisonics offers a convenient method of reproducing recordings obtained with a compact spherical microphone array since the processing can be done in the spherical harmonics domain [29].

For the Ambisonic reproduction, this paper uses a form of the spherical harmonics that employs real-valued trigonometric functions, which are more convenient for audio signals, based on the ambiX format convention [30]:

$$Y_n^m(\alpha, \phi) = \sqrt{\frac{(2 - \delta_{m0}) (n - |m|)!}{4\pi (n + |m|)!}} p_n^{|m|} (\sin \alpha) \cdot \begin{cases} \sin(|m|\phi) & \text{if } m < 0 \\ \cos(|m|\phi) & \text{if } m \geq 0 \end{cases}, \quad (3-8)$$

where $\alpha = \frac{\pi}{2} - \theta$ is the elevation angle relative to the horizontal plane, and $\delta$ is the Kronecker delta. Although not standard notation, the hat is used here to differentiate between the spherical harmonics in Eqn. (3-8) and Eqn. (3-3).

The signals to drive the individual loudspeakers for Ambisonic playback are calculated by:

$$g = D \hat{p}_n^m \quad (3-9)$$

where $g = (g_1, g_2, \ldots, g_Q)^T$ are the loudspeaker driving signals, $\hat{p}_n^m = (\hat{p}_0^0, \hat{p}_1^{-1}, \hat{p}_1^0, \ldots, \hat{p}_N^N)^T$ are the Ambisonics signals to be reproduced, which are encoded in the spherical harmonics domain, and $D$ is known as the Ambisonic decoder matrix. Several methods exist to design a decoder matrix for a given loudspeaker array, such as the traditional mode-matching or pseudoinverse method [31], all-round Ambisonic decoding [35], and energy-preserving decoding [69].
pseudoinverse method as implemented in the Ambisonics Decoder Toolbox [32, 33] was utilized in this study since it is well suited for nearly spherical loudspeaker distributions. Using this method, the basic decoder matrix is determined by taking the pseudoinverse of the following spherical harmonics matrix evaluated at the specific loudspeaker locations:

\[
D = \begin{bmatrix}
\hat{Y}_0^0(\alpha_{spk1}, \phi_{spk1}) & \hat{Y}_1^1(\alpha_{spk1}, \phi_{spk1}) & \cdots & \hat{Y}_N^N(\alpha_{spk1}, \phi_{spk1}) \\
\hat{Y}_0^0(\alpha_{spk2}, \phi_{spk2}) & \hat{Y}_1^1(\alpha_{spk2}, \phi_{spk2}) & \cdots & \hat{Y}_N^N(\alpha_{spk2}, \phi_{spk2}) \\
\vdots & \vdots & \ddots & \vdots \\
\hat{Y}_0^0(\alpha_{spkQ}, \phi_{spkQ}) & \hat{Y}_1^1(\alpha_{spkQ}, \phi_{spkQ}) & \cdots & \hat{Y}_N^N(\alpha_{spkQ}, \phi_{spkQ})
\end{bmatrix}^\dagger (3-10)
\]

The high frequency localization performance in Ambisonic reproductions can be improved by applying order-dependent gains to the decoder matrix, which is called max-\(r_c\) decoding. Applying these gains will effectively change the directivity pattern to reduce side lobes, which helps to preserve interaural level differences (ILDs). The implementation in Ref. [32] uses a phase-matched crossover to transition from basic decoding to max-\(r_c\) decoding at 400 Hz.

3.2 ROOM IMPULSE RESPONSE MEASUREMENTS

3.2.1 Measurement Hardware

Several pieces of hardware were utilized to capture the spatial room IRs. The primary hardware used consisted of an mh acoustics em32 Eigenmike spherical microphone array (Figure 3-2a), a Brüel & Kjær (B&K) Type 4100-D binaural mannequin (Figure 3-2b), and a Brüel & Kjær Type 4292 dodecahedron loudspeaker (Figure 3-2c). Additionally, a Crown XLS 2500 audio amplifier was used to drive the loudspeaker, and an RME Babyface was the audio interface used to send and receive signals.
The Eigenmike microphone array consists of 32 omnidirectional microphones mounted on a rigid spherical baffle with a diameter of 8.4 centimeters. The microphones sample the sphere according to the center of the faces of a truncated icosahedron, which is a nearly-uniform sampling scheme that preserves the orthogonality of the spherical harmonics up to 3\textsuperscript{rd} order. The em32 system contains the microphone array as well as the Eigenmike Interface Box (EMIB). The EMIB interfaces with a PC with a standard ASIO driver, which allows for synchronized IR measurements to be made with commercially available room acoustics software. For the excitation signal, the ADAT output of the EMIB was sent to the RME Babyface, which was used as a D/A converter. The output of the RME Babyface was connected to the amplifier to drive the dodecahedron loudspeaker.

In addition to spatial IRs obtained with the spherical microphone array, additional measurements were made using a Brüel & Kjær Type 4100-D head and torso simulator (HATS). The purpose of obtaining the binaural IRs was to be able to directly compare the original measured binaural IR with a binaural IR which was reproduced from the spherical array measurements (described in Section 3.3.2). For these binaural measurements, the RME Babyface was used as an external sound card to both drive the amplifier and record the two microphone channels. The dodecahedron loudspeaker was utilized in the same configuration as the spherical array measurements for these binaural measurements.
3.2.2 Room IR Measurements

IR measurements were obtained in the Peter Kiewit Concert Hall in Omaha, NE, a shoebox-shaped hall which opened in 2005. The hall has a volume of roughly 24,000 m$^3$ (850,000 ft$^3$) and 2,000 seats. The hall features variable absorption in the form of absorptive panels on the ceiling and walls. Measurements were made using the spherical microphone array in three different hall absorption settings: the most absorptive setting (Setting 1), the most reverberant setting (Setting 2), and a moderately reverberant setting (Setting 3), and in 10 to 11 receiver positions in the hall (Figure 3-3) depending on the hall setting. The mid-band reverberation times for these settings are 1.8 s, 2.8 s, and 2.4 s, respectively. For all of the measurements, the dodecahedron loudspeaker was placed in the center of the stage. Room IRs were obtained using the room acoustics software EASERA [63] with the multi-channel module. The excitation used was a 6-second logarithmic sine sweep, with eight averages and one pre-sweep.

![Figure 3-3: Receiver locations in the Peter Kiewit Concert Hall.](image-url)


3.3 STIMULUS REPRODUCTION

3.3.1 Auralization and Reproduction of Acoustic Sound-fields (AURAS) Loudspeaker Array

The Ambisonics reproduction in this study was conducted in the Auralization and Reproduction of Acoustic Sound-fields (AURAS) facility at Penn State University, shown in Figure 3-4a. The facility includes a loudspeaker array consisting of 30 two-way sealed-box loudspeakers. The loudspeakers feature a 4” (~10 cm) mid-bass driver and a 1” (~2.5 cm) fabric dome tweeter which are passively crossed over at 1.8 kHz. The loudspeakers are individually equalized from approximately 60 Hz to 20 kHz to account for magnitude and phase differences in the frequency response of each loudspeaker [36].

The 30 loudspeakers in the AURAS array are arranged in a nearly-spherical distribution as shown in Figure 3-4b. The majority of the loudspeakers are distributed over three rings: 8 loudspeakers located at α = -30°, 12 loudspeakers located at α = 0°, and 8 loudspeakers located at α = +30°. In each ring, the loudspeakers are distributed equally azimuthally with a loudspeaker located at φ = 0° directly in front of the listener. The average distance from the loudspeakers to the center of the array is r = 1.3 m. The remaining two loudspeakers are placed overhead at α = ±60°, φ = ±90°, and r = 0.57 m.

(a) (b)

*Figure 3-4: The AURAS loudspeaker array (a), and the distribution of the 30 loudspeakers in the array (b).*

Auralizations were generated for reproduction in the AURAS facility using the measured IRs described in Section 3.2. A block diagram for the stimulus generation is shown in Figure 3-5. Stimuli were rendered using the digital audio workstation software REAPER [70] and VST plugins from the Ambisonic Decoder Toolbox [32], and the ambiX and mcfx plug-in suites [71]. The
processing was applied to anechoic music files that were convolved with the IRs measured with
the spherical microphone array in MATLAB.

![Diagram of Ambisonic reproduction process]

Figure 3-5: Block diagram for Ambisonic reproduction

The first step in processing the microphone signals is to encode them into Ambisonic signals (i.e. spherical harmonic signals) using the Ambix format in Eqn. (3-8) with the ambiX_decoder plug-in with the Eigenmike preset. After Ambisonic encoding, radial filters are applied using the mcfx_convolver plug-in. These radial filters invert the plane wave modal coefficients in Eqn. (3-4) and are implemented as FIR filters. The radial filters, shown in Figure 3-6, are crossed over with linear phase filters at 50 Hz, 500 Hz, and 1.3 kHz for 1\textsuperscript{st}, 2\textsuperscript{nd}, and 3\textsuperscript{rd} order reproduction, respectively, and at n\textsuperscript{th} order, a (2n+1) correction factor is applied to preserve the pressure amplitude of the main lobe [72]. The crossover frequencies were chosen by measuring a plane wave with the spherical microphone array, performing a plane wave decomposition using Eqn. (3-7) with $c_n$ beamforming weights set to one, and determining the frequency in which the beam pattern begins to degrade for a given order. A random-incidence correction is included in these filters to equalize the frequency response of the spherical array’s microphone capsules, which have a high-frequency roll-off characteristic [37].
After radial filtering, the Ambisonic signals were decoded into the loudspeaker signals. The Ambisonic decoder was designed using the Ambisonic Decoder Toolbox [32] using the pseudoinverse method as described in Section 3.1.2. Finally, the decoded loudspeaker signals are equalized for the individual loudspeakers using the mcfx_convolver plug-in.

3.3.2 Validation

The performance of the Ambisonics reproduction in the AURAS facility was evaluated using both objective measurements and subjective listening tests. First, a plane wave oriented directly in front of the listener (\(\alpha = 0^\circ, \phi = 0^\circ\)) was encoded into Ambisonic signals, reproduced over the AURAS loudspeaker array, and measured with the spherical microphone array. A PWD was performed using the SOFiA toolbox [73] and the measured plane wave was compared to the results of a simulated plane wave by visual inspection. The measured plane wave matched up with the expected result at the appropriate crossover frequencies for 1\(^{st}\), 2\(^{nd}\) and 3\(^{rd}\) order, respectively, up to approximately 10 kHz. An example of the comparison at 2 kHz is shown in Figure 3-7, in which the plane wave exhibits the correct 3\(^{rd}\) order max-\(r_x\) beam pattern, which is expected for all frequencies above 1.3 kHz.
In addition to the measurements, informal listening tests were performed in which plane waves convolved with pink noise were encoded into 3rd order Ambisonics signals and panned in full 3D space. Listeners noted that the panning was smooth, even in between loudspeakers, and that they were able to localize sound in the correct directions except when sound was panned below $\alpha = -30^\circ$ elevation, where no loudspeakers were present.

As a perceptual validation of the Ambisonics reproduction, the binaural IRs described in Section 3.2.2 were compared with the reproduced spherical microphone array IRs. The binaural head was placed in the center of the AURAS array and the reproduced IRs were measured with the head. These IRs were convolved with anechoic music, and an ABX listening test was conducted to compare the reproduced binaural IR to the original measured IR. A monaural equalization was applied to the reproduced binaural IRs to match the average spectrum of the left plus right ears to the measured IRs while still maintaining the reproduced binaural cues. The ABX listening test was taken by six musicians, all with hearing thresholds at or below 15 dBHL for the 250 Hz through 8 kHz octave bands. The subjects reported that they had a difficult time distinguishing between the original IR and the reproduced IR, but were able to hear differences better between the stimuli by listening to a single note or small segment containing a small number of notes on repeat. For the subjects that were able to hear differences while listening to entire passages, these subjects needed a long time to make a decision (average of 30 seconds per trial). Based on the difficulty of this test, the time required to distinguish the differences, and the similarities noted by the listeners, the authors have concluded that the binaural Ambisonic reproductions are perceptually nearly indistinguishable from the original binaural recordings.
3.4 LISTENER ENVELOPMENT SUBJECTIVE LISTENING TEST 1: COMPARISON OF HALL SEATS AND ABSORPTION SETTINGS

A subjective study was carried out using the room IR measurements processed for Ambisonic reproduction in the AURAS facility in which participants were asked to rate stimuli in terms of perceived envelopment. The room IRs were convolved with an anechoic music excerpt, Bizet’s L’Arlesienne Suite No. 2: Menuet [74]. This piece contains a full orchestra including strings, winds, and timpani, and is played at a moderate tempo (72 beats per minute). The stimuli were presented in four sets of eight signals, which were presented to listeners in a randomized order. Set 1 contained a subset of eight IRs from the most absorptive setting, Set 2 contained a subset of eight IRs from the in-between setting, and Set 3 contained a subset of eight IRs from the most reverberant setting. Set 4 contained a mixture of IRs from the three aforementioned settings that all had similar $L_I$ values.

During the subjective test, each listener was seated in the center of the loudspeaker array and was able to listen to the different stimuli via a graphical user interface (GUI) implemented in Max 7 [75]. The GUI presented the participants with a screen with individual buttons for all of the stimuli and the subjects were able to instantaneously switch between each of the stimuli without restarting the musical motif. The GUI also enabled listeners to limit the motif to a specific time segment of the passage. Each subject was asked to rate how enveloped they felt by the sound field on a scale from 0 (not at all enveloped) to 100 (completely enveloped).

Before beginning the test, participants completed a short training period. The training began with a tutorial explaining the GUI, with explicit instructions to focus only on LEV, the sense of being surrounded by or immersed in the sound, and to ignore all other aspects of the sound field, including apparent source width. Following the tutorial, participants performed a training session to learn how the GUI worked with a reduced stimuli set of four audio files. After the training session, participants were then given a practice set containing a full set of eight stimuli. However, the participants were instructed that this was the first set in the test, but in actuality this data was not used in the analysis.

The listening test was conducted with 15 participants (6 male and 9 female). All subjects were required to have measured hearing thresholds at or below 15 dBHL from 250 Hz to 8000 Hz, since research has shown that in critical listening tests subjects with near-normal hearing
thresholds provide more consistent ratings [76]. Additionally, all subjects were required to have a minimum of 5 years of formal musical training and be musically active at the time of the study (i.e. performing in an ensemble and/or taking private music instruction). This requirement was imposed because musicians have been shown to learn listening test procedures more quickly and to give more consistent responses compared to non-musicians [77] [78]. The average age of the participants in the subjective test was 26 years old, with an average of 10 years of formal musical training.

3.4.1 Subjective Listening Test 1 Results

Each set was analyzed separately using a one-way repeated-measures analysis of variance (ANOVA) [65]. The null hypothesis of this ANOVA test was that the mean LEV ratings of all of the stimuli within a set are identical and a significant p-value would indicate that at least one pair of the stimuli within a set have statistically different mean LEV ratings. The results of the statistical analysis yielded significant results for Sets 1 through 3 (p < 0.001, p < 0.001, and p = 0.014, respectively), which indicates that these sets contain significant differences in the mean LEV ratings of at least one pair of stimuli within each set. The result for Set 4, however, was found to be insignificant on a 95% confidence interval (p = 0.071). Therefore, none of the pairs of stimuli in Set 4 were found to contain differences in mean LEV ratings. One possibility for the insignificant results is that the number of subjects in the study was too low to detect such small differences between the stimuli in this set, which all had similar L1 values, leading to low statistical power. The mean LEV ratings from the four test sets are shown in Figure 3-8.
Figure 3-8: Mean LEV ratings for the four test sets. Error bars depict standard errors.

For the three sets with significant p-values, pairwise t-tests were conducted to determine the individual pairs with significant differences. Within Sets 1 and 2, pairs were identified in which the LEV ratings were significantly different but values of \( L_I \) were identical. Conversely, within the same sets, pairs were found in which the \( L_I \) values differed by up to 1.3 dB, but the LEV ratings were found to be similar. In order to examine the relationship between LEV ratings and the 3D late sound field from the stimuli in Sets 1 through 3 in more detail, the measurements were analyzed using beamformed IRs (Eqn. (3-7)), which is described in in Section 3.5.

The LEV ratings obtained in each set were fit to a one-way regression model using different room acoustics metrics as a predictor for LEV: EDT, T30, C80, G, \( G_{\text{Late}} \), \( L_I \), LLF, SBT, and front/back ratio. (See Appendix A for the details about how SBT and front/back ratio were

---

\(^2\) Note that the same stimulus between two different sets may have a different absolute LEV rating, since the stimuli are rated relative to other stimuli within the set.
calculated from spherical microphone array measurements.) Set 1 contained the largest differences in LEV ratings amongst the eight stimuli. The correlation coefficients (R) and associated p-values for each of the regression models are given for Set 1 in Table 3-1. The metrics that show the highest correlation with LEV in multiple octave bands are the metrics that are related to overall level: strength (G), late strength (Gₐₗₑₜ), and late lateral energy level (Lⱼ). The metric with the highest correlation was Gₐₗₑₜ, which had a correlation of r = 0.88 (p < 0.01) in the 500 Hz octave band. It should also be noted that Lⱼ was only significantly correlated with LEV at 500 Hz and above, with correlation coefficients ranging from r = 0.76 to r = 0.8, although this metric was originally defined to be related to LEV from 125 Hz to 1 kHz [10].

*Table 3-1: Correlation coefficients for different metrics as LEV predictors for Set 1.*

<table>
<thead>
<tr>
<th>Metric</th>
<th>Significant Octave Bands (Hz)</th>
<th>Correlation Coefficient, R (by octave band)</th>
<th>p-Value (by octave band)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EDT</td>
<td>None</td>
<td>N/A</td>
<td>&gt;0.05</td>
</tr>
<tr>
<td>T30</td>
<td>125</td>
<td>-0.80</td>
<td>0.017</td>
</tr>
<tr>
<td>C80</td>
<td>125</td>
<td>-0.78</td>
<td>0.022</td>
</tr>
<tr>
<td>G</td>
<td>500, 1k, 4k</td>
<td>0.73, 0.79, 0.76</td>
<td>0.040, 0.020, 0.028</td>
</tr>
<tr>
<td>Gₐₗₑₜ</td>
<td>500, 2k</td>
<td>0.88, 0.77</td>
<td>0.004, 0.024</td>
</tr>
<tr>
<td>Lⱼ</td>
<td>500, 1k, 2k, 4k</td>
<td>0.77, 0.76, 0.79, 0.8</td>
<td>0.026, 0.027, 0.019, 0.17</td>
</tr>
<tr>
<td>LLF</td>
<td>None</td>
<td>N/A</td>
<td>&gt;0.05</td>
</tr>
<tr>
<td>SBTS</td>
<td>None</td>
<td>N/A</td>
<td>&gt;0.05</td>
</tr>
<tr>
<td>FBR</td>
<td>250</td>
<td>-0.76</td>
<td>0.029</td>
</tr>
<tr>
<td>FBRₐₗₑₜ</td>
<td>None</td>
<td>N/A</td>
<td>&gt;0.05</td>
</tr>
</tbody>
</table>
3.5 BEAMFORMING ANALYSIS OF THE MEASURED IMPULSE RESPONSES

The spatial IRs measured with the spherical microphone array were analyzed objectively using beamforming techniques in order to investigate the relationship between the results of the subjective study to the physical sound field. Directional IRs were generated by applying a spatial Fourier transform as in Eqn. (3-6), applying the radial filters shown in Figure 3-6, and beamforming via Eqn. (3-7) using look-directions spaced every 3 degrees in azimuth and elevation. This process results in a 120 by 60 (azimuth by elevation) grid of IRs. Two sets of grids were generated: one grid using PWD coefficients for regions where 2\textsuperscript{nd} and 3\textsuperscript{rd} order spherical harmonics components are used, and one grid using a cardioid-type pattern for low frequencies in which only the 1\textsuperscript{st} order components are usable. A cardioid-type pattern was used for low frequencies since the 1\textsuperscript{st} order plane wave pattern has a large rear lobe, which can lead to misinterpretations of the energy plots as the energy appears to originate from the wrong direction.

The directional IRs were further processed to investigate the timing and direction of the energy in the IRs as a function of frequency. First, a 5 ms-wide rectangular time window was applied to the IRs every 5 milliseconds (i.e. 0 to 5 milliseconds, 5 to 10 milliseconds, etc.). The time windowed IRs were then filtered into octave bands with center frequencies from 125 Hz to 4 kHz, and the energy in each octave band was summed. Using this analysis approach, the energy in different time windows can be plotted on a grid as a function of angle for different octave bands with a resolution of 5 milliseconds. For example, to investigate the early sound field, the energy contained in the 5 millisecond time segments from 0 to 80 milliseconds can be summed together and plotted on a grid.

For each of the IRs, the late energy (80 milliseconds onward) was investigated, and instances were found where the energy in the late sound field did not correlate with the subjective envelopment ratings. For example, the LEV ratings in Set 2 of R3 (L\textsubscript{E} = -1.4 dB) and R8 (L\textsubscript{E} = -0.1) were nearly identical (73 and 72, respectively), but the late sound fields are very different, as shown in Figure 3-9. In Figure 3-9 and all subsequent plots the 1 kHz octave band results are shown, although the trends are observed broadband (500 Hz – 4k Hz). The late energy in R3 is concentrated in the front of the listener and overhead with lower side energy, whereas the
energy in R8 appears more diffuse with the most energy coming from behind the listener. Differences in the PWD for the late arriving lateral energy are on the order of 3 dB.

Conversely, instances were found in which the stimuli had significantly different LEV ratings with similar $L_j$ values. The stimuli in one example pair found in Set 1 (R9 and R10) have nearly identical values of $L_j$ (-3.3 dB and -3.6 dB, respectively) but very different LEV ratings (66 and 48, respectively, $p = 0.020$), which can be seen in Figure 3-10. In these receiver positions, the late energy distributions appear to be very similar, and energy differences in the lateral energy in the PWD is on the order of 1-2 dB, which are much lower than the differences seen in the previous pair. The results from these pairs indicate that the directional distribution of late arriving energy does not completely predict the sense of LEV.

**Figure 3-9:** Comparison of the late sound field at 1 kHz (80 ms to $\infty$) between receiver positions with similar LEV ratings (R3 and R8 from Set 2), with 0° azimuth pointing toward the stage and 0° elevation pointing straight up. Sound pressure level is shown ranging from -10 dB to 0 dB, where 0 dB is the maximum level of both sound fields (overall level differences are maintained between the top and bottom plots). The images on the left show the energy distributions over spheres, while the images on the right show the same information, but flattened onto a 2-D plot (similar to an unraveled map.) Energy at R3 is concentrated toward the front (R3 is underneath a balcony), whereas energy at R8 is more evenly distributed throughout the sphere (R8 is in the top balcony).
Set 1: Receiver 9 (LEV Rating = 66)

Set 1: Receiver 10 (LEV Rating = 48)

Figure 3-10: Comparison of the late sound field at 1 kHz between receiver positions with different LEV ratings (R9 and R10 from Set 1), with 0° azimuth pointing toward the stage and 0° elevation pointing straight up. Sound pressure level is shown ranging from -10 dB to 0 dB, where 0 dB is the maximum level of both sound fields (overall level differences are maintained between the top and bottom plots). The spatial distribution of late energy is similar between the two receivers yet the LEV ratings of these stimuli were significantly different.

Because of the anomalies found in the late sound fields, different time windows were explored to find trends in different parts of the IRs that correlated with the LEV ratings. These additional time windows included the early sound (0 to 80 milliseconds), late IRs with different crossover points (e.g. 40 milliseconds onward or 200 milliseconds onward), and portions in the middle of the IRs (e.g. 50 milliseconds to 200 milliseconds). Using this exploratory approach, a time window in the middle of the IR from 70 to 100 milliseconds was found to agree with the LEV ratings. In this time window, R3 and R8 in Set 2 have a very similar distribution of lateral, back, and overhead energy, which can be seen in Figure 3-11. The frontal energy in this time window is approximately 4 dB greater in R3, but the frontal energy is assumed to not significantly impact LEV [10] [13]. The energy distribution in this time window could be related to the similarities in LEV ratings. For R9 and R10 in Set 1 in this time window, differences can be seen in that the rear energy is roughly 3 dB higher behind the listener in R9 than in R10, shown in Figure 3-12, which
could be related to the differences in LEV ratings. In order to confirm the trends that were found using this time window, a second subjective study was conducted using modified versions of the IRs, described in Section 3.6.

Figure 3-11: Comparison of the sound field from 70 to 100 ms at 1 kHz between receiver positions with similar LEV ratings (R3 and R8 from Set 2), with 0° azimuth pointing toward the stage and 0° elevation pointing straight up. Sound pressure level is shown ranging from -10 dB to 0 dB, where 0 dB is the maximum level of both sound fields (overall level differences are maintained between the top and bottom plots). The energy at both receiver positions have a similar level and distribution in terms of lateral, behind, and overhead sound. The frontal energy does differ by 3 dB between the pair, but it is assumed that energy arriving from the front does not influence LEV.
3.6 LISTENER ENVELOPMENT SUBJECTIVE LISTENING TEST 2: LISTENING TEST USING HYBRID IMPULSE RESPONSES

In order to investigate the influence of arrival time on the energy in the spatial IR, a follow-up subjective study was conducted using modified IRs created by mixing different portions of two IRs together with a variable crossover time. Two types of hybrid IRs were created for this study: (1) combinations of a pair of IRs that were rated to have the highest and lowest amounts of LEV from Set 1 and (2) combinations of the IR found to have the highest LEV with a modified version of that IR in which the spatial dependence was collapsed into a single direction directly in front of the listener. The objective of the listening tests using the first type of hybrid IRs was to determine at which point in time crossing over from a highly enveloping IR to an unenveloping IR or vice versa impacts the perception of envelopment. The listening test using the second type of hybrid IRs was used to study the same effect in the case where all spatial information is removed from portions of the IR.
For the first type of hybrid IRs, two stimuli were used from the previous listening test. The stimulus with the highest LEV rating from Set 1, which contained stimuli from the most absorptive hall setting (Setting 1), was at R3, while the stimulus with the lowest LEV rating was at R10. These IRs will be referred to as the high LEV (HLEV) IR and the low LEV (LLEV) IR, respectively. Hybrid IRs were generated that contained the early part of the HLEV IR and the late part of the LLEV IR using crossover times ranging from 40 ms to 140 ms – denoted as the HLEV/LLEV hybrid IR. Similarly, hybrid IRs were also created that contained the early part of the LLEV IR (Setting 1, R10) and the late part of the HLEV IR – denoted as LLEV/HLEV hybrid.

For the second set of hybrid IRs, a more extreme modification of the IRs was utilized in which the spatial aspects of the sound field were completely removed from portions of the HLEV IR (Setting 1, R3). A monaural IR was generated by extracting the omnidirectional component of the HLEV IR and sending it to the loudspeaker located directly in front of the listener, which collapsed the full 3D sound field into the single direction at $\alpha = \phi = 0^\circ$. Hybrid IRs were then generated containing the early part of the HLEV IR and the late part of the monaural IR – denoted as the 3D/mono hybrid IR, and vice-versa – denoted as the mono/3D hybrid IR, with crossover times ranging from 40 ms to 120 ms, respectively.

A 2.5 ms half-Hann window was used to crossover between the early part and the late part of the IRs, as depicted in Figure 3-13. This window ensured that there was no audible click in the transition period, while being short enough to not have a noticeable transition. The stimuli were generated by convolving the hybrid IRs with the same anechoic music excerpt used in the first study, Bizet’s L’Arlesienne Suite No. 2: Menuet [74].
Figure 3-13: Half Hann 2.5 millisecond time windows used to mix two IRs (left), and example resulting hybrid IR (right). The early window and corresponding IR are shown in solid blue, and the late window and corresponding IR are shown in dotted and solid green, respectively.

3.6.1 Hybrid IR subjective test design

During the subjective test, each listener was placed in the center of the loudspeaker array and was able to listen to the different stimuli via instantaneous switching using the GUI discussed in Section 3.4 with modifications to include a different number of stimuli for the test cases described below. All subjects were given eight sets of stimuli presented in a random order. Each subject was asked to rate how enveloped they felt by the sound field on a scale from 0 (not at all enveloped) to 100 (completely enveloped). The listening test was conducted with 19 participants (10 male, 9 female). Two subjects out of the 19 participated in the previous listening test. All subjects had measured hearing thresholds at or below 15 dBHL from 250 Hz to 8000 Hz, were required to have a minimum of 5 years of formal music training, and were required to be musically active. The average age of the participants was 23 years old, with an average of 8 years of formal musical training.

Six of the test sets contained the HLEV/LLEV and LLEV/HLEV hybrid stimuli. Each of the six sets contained four stimuli presented simultaneously in a random order: stimuli generated with the complete HLEV IR, the complete LLEV IR, the HLEV/LLEV hybrid IR, and the LLEV/HLEV hybrid IR. All of the six sets contained these four stimuli, each having a different crossover point between
the early and late sound for the hybrid IRs, which varied from 40 ms to 140 ms in 20 ms increments.

The remaining two sets contained the 3D/m mono hybrid IRs, and the mono/3D hybrid IRs, respectively. A MUSHRA (MUltiple Stimuli with Hidden Reference and Anchor) style listening test was implemented for these two sets [79]. The reference used in the sets was the full 3D IR, and the anchor was the full monaural IR. Listeners were presented with the reference, which was labeled, and seven stimuli, two of which were the hidden anchor and the hidden reference. The remaining five stimuli had crossover times ranging from 40 ms to 120 ms in 20 ms increments. Listeners were instructed to treat the reference as having a rating of 100 (completely enveloped).

3.6.2 Results for hybrid IRs containing portions of an unenveloping IR and an enveloping IR

The six sets of IRs containing the HLEV/LLEV hybrid stimuli were combined into two data sets: one set containing the HLEV/LLEV ratings and one set containing the LLEV/HLEV ratings. These two sets were analyzed separately using one-way repeated measures ANOVA tests [65]. Pairwise t-tests were run using Tukey family error-rate corrections [65]. The results using the IRs from the early part of R3 and the late part of R10, shown in Figure 3-14, suggest that the 40 ms and 60 ms crossover times were nearly indistinguishable from the original R10 IR in terms of LEV. Additionally, 140 ms crossover time is nearly indistinguishable from the R3 IR in terms of LEV. The in-between crossover times have LEV ratings in between R3 and R10. These results indicate that in this set most of the auditory cues related to the perception of LEV for these two stimuli is contained in the time limits of 80 and 120 ms.
Figure 3-14: Modified IRs using the early part of R10 (highly unenveloping) and the late part of R3 (highly enveloping) with crossover times ranging from 40 ms (highly enveloping) to 140 ms (highly unenveloping).

An opposite trend was observed in the hybrid IRs containing the early part of R10 and the late part of R3, which can be seen in Figure 3-15. Although the 40 ms crossover time stimulus was very similar to the original R3 IR, the LEV rating of the stimulus was significantly lower than that of the original. Here, the 120 and 140 ms crossover times were nearly indistinguishable from the original R10 IR in terms of LEV. The 60 ms to 100 ms crossover times had an LEV rating in between the LEV ratings of R3 and R10, which indicates that this range contains information important to the perception of LEV. In this listening test, more than half of the total change in envelopment occurs within the first 80 ms of crossover times. These results suggest that the early part of the IR before 80 ms influences the perception of LEV. Moreover, a prediction of LEV which only includes energy later than 80 ms may not be sufficient for this pair of stimuli.
Figure 3-15: Modified IRs using the early part of R3 (highly enveloping) and the late part of R10 (highly unenveloping) with crossover times ranging from 40 ms (highly enveloping) to 140 ms (highly unenveloping).

3.6.3 Results for hybrid IRs with removed spatial components

The two sets containing the 3D/mono and mono/3D hybrid stimuli were analyzed separately using one-way repeated measures ANOVA tests [65]. The results from the 3D/mono listening test can be seen in Figure 3-16. The extreme contrast between the reference and the anchor resulted in the subjects utilizing much more of the entire scale from 0 to 100 than in the previous listening tests (mean range of 81 versus 25). By replacing just the first 40 ms of the mono IR with the spatial IR, there is already a large increase in the average LEV rating. Increasing the crossover time increases the average LEV rating, and at the 120 ms crossover point, the average LEV rating is not quite as high as the originally fully 3D IR. These results again illustrate that the early part of the IR does indeed have an effect on the perception of LEV.
Results from the mono/3D hybrid listening test have the opposite trend as the results from the 3D/mono test set, as seen in Figure 3-17. Collapsing just the first 40 ms of the IR into the front of the listener causes a large degradation in average LEV rating, from 92 to 73. Increasing the crossover times (collapsing more and more of the IR in time) results in a decrease in LEV ratings. The 120 ms crossover time did not have as low of an LEV rating as the full mono IR, which could be because even with the 120 ms crossover time, it is still easy to distinguish the spatial part of the IR from the collapsed portion.
3.7 CONCLUSIONS

In this study, using a 32-element spherical microphone array, spatial room IRs were obtained in a 2,000 seat dedicated concert hall with a volume of 24,000 m$^3$ (850,000 ft$^3$) (Peter Kiewit Concert Hall in Omaha, NE, USA). The hall features variable acoustics with mid-frequency reverberation times ranging from 1.8 s to 2.8 s. The IRs were analyzed using beamforming techniques and processed for 3$^\text{rd}$ order Ambisonic reproduction in order to conduct listening tests. The Ambisonics playback system was validated both objectively and perceptually. LEV scores from the subjective listening tests were compared to several objective measures and to the energy in the beamformed directional IRs in different time windows.

To determine whether or not existing objective metrics were correlated with the perception of LEV, linear regressions were conducted with the LEV ratings as the response variable and individual metrics as the predictors. The predictors found with the highest correlations were all functions of sound level: strength (G), late strength ($G_{\text{Late}}$), and late lateral energy level ($L_l$). This finding indicates that throughout this hall, the sense of envelopment is highly dependent on level. Metrics having to do with spatial aspects of the room, but not overall level, including lateral fraction ($L_F$), late lateral fraction (LLF), front-back ratio (FBR), and spatially balanced center time (SBT$_S$), were found to have little to no correlation with the LEV ratings. However, since all of the stimuli from this study were in the same hall, it could be the case that in halls of different sizes and shapes, LEV could be better correlated with these metrics.

In comparing the spatial distribution of the energy in the IRs to the LEV ratings, pairs of stimuli emerged with statistically different mean LEV ratings, but similar late sound fields and vice versa. An exploratory look into different time windows found that there were trends in the energy distribution in an earlier time window, specifically from 70 ms to 100 ms.

To investigate specific time windows in the IRs, hybrid IRs were generated using a mix of an enveloping (HLEV) IR and an unenveloping (LLEV) IR, as well as hybrid IRs using a mix of the original fully spatial IR (3D) and an IR in which the energy was collapsed to a single direction in front of the listener (mono). For the hybrid IR sets that contained the early part of the HLEV IR and late part of the LLEV IR (HLEV/LLEV), as well as the sets that contained the early part of the LLEV IR and the late part of the HLEV IR (LLEV/HLEV), it was found that a crossover time after 120 ms had very little effect on the envelopment. It was also found that a crossover time
between 80 and 120 ms in the HLEV/LLEV hybrids, or 60 and 100 ms in the LLEV/HLEV hybrids, resulted in an LEV rating halfway in between the original enveloping and unenveloping IRs. This finding was in agreement with the spatial energy distribution found in the beamforming analysis of the IRs. This finding suggests that this time region contains information important to the perception of LEV and that the early part of the impulse response should be taken into account in metrics predicting LEV. Hybrid listening tests containing 3D/mono and mono/3D hybrids confirmed that changes in the early part of the IR impacts the LEV ratings.

One important note is that these time crossovers are only appropriate for this hall and cannot yet be extrapolated to the general case. The late part of the IR was relatively constant throughout this hall, which could be a reason that modifications to the IRs after 120 ms had a smaller impact on LEV ratings. Similar tests should be conducted using measured spatial room IRs from a number of halls with a range of shapes and sizes to identify the relative importance of different time segments and arrival direction of the energy in the IR in creating the sense of LEV.

3.8 ACKNOWLEDGEMENT

The authors wish to acknowledge Matthew Neal for his assistance with the project and the AURAS facility; Ed Hurd at the Peter Kiewit Concert Hall; and Dr. Lily Wang, Matthew Blevins, Laura Brill, Hyun Hong, Joonhee Lee, and Zhao Ellen Peng of University of Nebraska-Lincoln for their assistance in obtaining the IR measurements. Approval for human subjects testing was obtained from Penn State’s Institutional Review Board (IRB #41733). This work was sponsored by the National Science Foundation (NSF) award #1302741.

3.9 APPENDIX A: CALCULATING METRICS FROM SPHERICAL MICROPHONE ARRAY IR MEASUREMENTS

The room acoustics metrics were calculated from the 32-channel spherical microphone array IR measurements, which were taken using an Eigenmike. For the standard metrics outlined in ISO 3382 [17], along with any metrics involving an omnidirectional IR and a figure-8 IR, the zeroth and first order components were extracted from the spatial IR measurements to compute the metrics as shown in Ref. [37]. For the additional metrics evaluated, specifically SBT, [24] and FBR [13], the methods to compute these metrics from IR measurements are not well defined in the literature. The ambiguity results from the metrics being developed based on room simulations in which all reflections come from discrete angles where loudspeakers were placed, and not from
physical measurements. In measurements obtained in rooms, however, reflections can come from any angle, and the measurement microphones have a directivity pattern with some finite beam width. The authors have made their best attempt at obtaining these metrics from measured IRs.

3.9.1 Front Back Ratio (FBR)
The front back ratio, FBR, is defined as:

\[
FBR = 10 \log \left( \frac{E_f}{E_b} \right),
\]

(A1)

where \( E_f \) and \( E_b \) is the energy in the IR in the front half of the horizontal plane and the back half of the horizontal plane, respectively [22]. For this study, the FBR was calculated in each octave band from 125 Hz to 4 kHz separately. Additionally, the FBR was calculated for the total impulse response as a function of time, as well as for only the late part from 80 ms onward.

To calculate this metric, a beamformed first-order cardioid pattern was used to obtain energy from the front (pointed directly at the loudspeaker at \( \theta = \pi/2, \phi = 0 \)) and energy from the back (pointed directly in the back at \( \theta = \pi/2, \phi = \pi \)). These beampatterns are a reasonable approximation for the energy in the front half of the horizontal plane and the energy in the back half of the horizontal plane, respectively.

3.9.2 Spatially Balanced Center Time (SBT)<sub>s</sub>
To calculate the spatially balance center time, SBT<sub>s</sub>, a PWD was performed on the spatial IRs with the look direction oriented in 16 equally spaced angles in the horizontal plane, which were the angles used to develop the metric [24] based on the results of their subjective study. Three separate PWDs were performed up to order \( N = 3 \) as a function of frequency: 1<sup>st</sup> order from 50 Hz to 500 Hz, 2<sup>nd</sup> order from 500 Hz to 1.3 kHz, and 3<sup>rd</sup> order from 1.3 kHz to 8 kHz. The SBT<sub>s</sub> was calculated separately for each order.

For each of the 16 directionally beamformed IRs, first a center time \( T_{Si} \) is calculated,

\[
T_{Si} = \frac{\int_{0}^{\infty} t \cdot p_i^2(t) dt}{\int_{0}^{\infty} p_i^2(t) dt},
\]

(A2)
where \( p_i(t) \) is the pressure beamformed in the specified direction, and \( p(t) \) is the omnidirectional pressure. To account for the crosstalk between each of the beamformed IRs, the individual center times are scaled by the sum of the center times,

\[
T_{\hat{S}i} = T_{Si} \cdot \frac{T_s}{\sum_i T_{Si}}. \tag{A3}
\]

The center times are weighted by the arrival direction such that energy directly in front or behind the listener is weighted by 0.5, and lateral energy is weighted by 1,

\[
a_i = T_{\hat{S}i} \cdot \frac{1 + |\sin \phi_i|}{2}, \tag{A4}
\]

where \( \phi_i \) is the look direction of the beamformed directional impulse response. SBT\( \_s \) is calculated by weighting the \( a_i \) terms by the contributions from the other directions and the sine of the angle in between the other dimensions,

\[
SBT_s = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i \cdot a_j \cdot \sin \phi_{ij}, \tag{A5}
\]

where \( \phi_{ij} \) is the angle in between look direction \( i \) and look direction \( j \).

The SBT\( \_s \) obtained from the spatial IR measurements were validated using simulated data recreating Test 3 from Ref. [24]. A simulated sound field was generated with a 2 second mid-band reverberation time and all energy coming from the directions specified by the test case. While the absolute level of SBT\( \_s \) was different than the calculated results reported in the article, which was expected since the simulated sound fields were not identical, the shape of the curve matched up identically, indicating that the metric obtained using the spatial IR measurements is performing as expected.
Chapter 4: A New Metric to Predict Listener Envelopment based on Spatial Impulse Response Measurements and Subjective Listening Tests

This chapter will discuss spatial room IR measurements obtained in seven different halls, a subjective study comparing the LEV ratings of the halls, and will outline the development of a new metric to predict LEV based on spherical microphone array measurements. This text is formatted as a manuscript which will be submitted for review and publication in a peer-reviewed journal.
A New Metric to Predict Listener Envelopment based on Spatial Impulse Response Measurements and Subjective Listening Tests

ABSTRACT

The objective of this work was to create a new metric to predict listener envelopment (LEV), the sense of being surrounded by the sound field. Spatial impulse responses (IRs) were obtained in seven halls of various sizes and shapes using a 32-channel spherical microphone array. The IRs were convolved with anechoic music and processed for 3rd order Ambisonics reproductions for listening tests in which subjects rated LEV of the stimuli. The listening tests included stimuli that were presented as-measured, which included level differences, stimuli that were equalized for level differences, and hybrid stimuli generated by combining portions of enveloping IRs and unenveloping IRs. The IRs were analyzed objectively using a 3rd order plane wave decomposition (PWD) beamformer. A new metric was developed by integrating energy from the PWDs as a function of frequency, azimuthal angle, elevation angle, and time, and adjusting the integration limits to maximize correlation between integrated energy and LEV ratings. The correlation was maximized by integrating from 60 ms to 400 ms, rejecting sound from the front ±20° in azimuth, and rejecting sound from ±70° in azimuth behind the listener. In one test set, the conventional metric yielded $R^2=0.77$ whereas the proposed metric yielded $R^2=0.94$.

4.1 INTRODUCTION

Listener envelopment (LEV), the sense of being surrounded by the sound field, is an important perceptual attribute of a performing arts space that has been shown to be correlated with the overall preference of a hall [1] [2]. LEV has been found to be a distinct component of the spatial impression of a hall, separate from the sense of apparent source width (ASW) [9]. Further work on spatial impression concluded that the ASW is related to the early lateral reflections in the hall prior to 80 ms in the impulse response (IR) while LEV is related to late lateral energy arriving after 80 ms [67].

A study conducted by Bradley and Soulodre [10] varied parameters of the late sound field including reverberation time (T30), the early-to-late energy ratio (clarity index, C80), the level of the late sound field ($G_{Late}$), and the angular distribution of the late sound field. The properties of the sound field that were shown to have the highest correlation between these parameters and
LEV were the angular distribution of the sound and the late sound level. These findings led to the development of a metric to predict LEV called late lateral energy level, $L_j$ (prior notations $GLL$, $LG$, and $LG_{80}^{\infty}$):

$$L_j = 10 \log_{10} \left[ \frac{\int_{80ms}^{\infty} p_L^2(t) dt}{\int_{0}^{\infty} p_{10}^2(t) dt} \right] \text{[dB]},$$

(4-1)

where $p_L(t)$ is the room IR measured with a figure-of-eight microphone with the null plane oriented toward the source, and $p_{10}(t)$ is the IR of the sound source normalized at a distance of 10 meters in a free field. Currently, $L_j$ is the most widely accepted metric to predict LEV and is included in Appendix A of the ISO-3382 standard [43].

One of the assumptions in Bradley and Soulodre's experimental design is that the early sound field does not impact LEV, and thus, the simulated early reflections in their listening test stimuli were held constant. However, several studies suggest that the early sound field does affect the perception of LEV. One study found that early reflections from the ceiling affected LEV [12] and another study notes that LEV seemed to be affected by the early sound field based on listening tests where the stimuli were reproduced over loudspeakers [24]. A study using binaural IRs, where portions of the left and right channels were cross-mixed into the other channel, found that manipulating the early part of the IR impacted LEV [16].

In a previous study by the authors (see Chapter 3), LEV was studied using spatial room IRs measured with a 32-element spherical microphone array in several seating positions and variable absorption configurations in a single hall. IRs were obtained in the hall at 10-11 receiver locations in three different absorption settings with mid-frequency T30 values ranging from 1.8 seconds to 2.8 seconds. Stimuli were generated and presented to listeners over a 3D 30-channel loudspeaker array such that they were able to instantly switch between different listener positions. Hybrid stimuli were also generated which contained portions of a highly enveloping IR and a highly unenveloping IR. The results showed that with the hybrid stimuli, LEV was more correlated with the early sound field than with the late sound field. Additionally, the metrics found to have the highest correlation to LEV were those that included a level term. However, it must be noted that all of the stimuli came from a single 2000-seat concert hall.
The purpose of this study was to develop a metric that can predict LEV more accurately than the current state-of-the-art metrics. Room IRs were obtained in seven different performing arts spaces of different sizes and shapes using a 32-element spherical microphone array. The IRs were convolved with a single orchestral music excerpt and processed for 3rd order Ambisonic playback over a 30-channel loudspeaker array, which was used in the four listening test sets. The first listening set contained IRs from each hall measured with the same source-receiver distance. The second test set contained the same IRs from the first test set, with a level equalization such that each stimulus had the same overall sound pressure level (dBA). The third and fourth test sets were hybrid IRs composed of time portions from both enveloping and unenveloping IRs with different time crossovers. The spatial IRs were also analyzed using plane wave decomposition (PWD). This objective analysis combined with the listening test results were used to develop a new metric to predict LEV.

4.1.1 Spherical Microphone Array Beamforming

Measurements obtained from a spherical microphone array can be transformed into the spherical harmonics domain:

\[
\overline{P}_{nm}(f) = \frac{1}{b_n(f)} \frac{4\pi}{S} \sum_{s=1}^{S} P_s(f) Y_{nm}(\theta_s, \phi_s),
\]

where \(\overline{P}_{nm}\) is the spatial Fourier coefficient of order \(n\) and degree \(m\), \(P_s\) is the pressure measured at microphone \(s\) in the frequency domain, \(b_n(f)\) are modal coefficients dependent on the geometry of the array, \((\theta_s, \phi_s)\) are the elevation and azimuth locations of microphone \(s\), and \(Y_{nm}\) are spherical harmonics of order \(n\) and degree \(m\) evaluated at the microphone locations:

\[
Y_{nm}(\theta, \phi) = \sqrt{\frac{(2n+1)(n-m)!}{4\pi(n+m)!}} P_n^m(\cos\theta) e^{im\phi}.
\]

The measurements can then be beamformed by weighting and summing the spherical harmonic components:
\[ P(\theta_l, \phi_l) = \sum_{n=0}^{N} c_n \sum_{m=-n}^{n} \tilde{P}_{nm}(ka)Y_n^m(\theta_l, \phi_l), \]  

(4-4)

where \((\theta_l, \phi_l)\) is the look direction of the beam and \(c_n\) is an order-dependent weighting that defines the beam pattern. For a plane wave decomposition (PWD), the \(c_n\) weights are set to one. Please see Chapter 1 Section 1.2 and Chapter 3 Section 3.1.1 for a detailed discussion of the beamforming process.

### 4.1.2 Ambisonics Reproduction

Ambisonics was used as a method to generate the sound fields for listening tests in this study [27] [28]. Ambisonics utilizes a spherical harmonic representation of the sound field and aims to generate that sound field at the center of a loudspeaker array. The Ambisonics implementation utilized in this study was the Ambisonics Decoder Toolbox [32] [33]. The Ambisonic decoder reproduces 3rd order spherical harmonic components, and includes nearfield compensation [80] and max-\(r_e\) decoding above 400 Hz [81].

The Ambisonic reproductions for listening tests were conducted in the Auralization and Reproduction of Acoustic Sound-fields (AURAS) facility at Penn State University, as shown in Figure 4-1. The AURAS facility is comprised of a 30-loudspeaker array located in an anechoic chamber, which has free-field characteristics down to the 200 Hz one-third octave band. The loudspeakers are located in three rings: 8 loudspeakers at 30° below the horizontal plane, 12 loudspeakers at the horizontal plane, and 8 loudspeakers at 30° above the horizontal plane. The loudspeakers are distributed equally azimuthally. The average distance from the loudspeakers to the center of the array is \(r = 1.3\) m. Two additional loudspeakers are placed overhead at 60° above the horizontal plane at ± 90° in azimuth and \(r = 0.57\) m. For more information about the AURAS facility, please see Ref. [36]. For a detailed discussion of Ambisonics, see Chapter 1 Section 1.3 and Chapter 3 Section 3.1.2.
4.2 SPATIAL ROOM IR MEASUREMENTS

Room IRs were obtained using a spherical microphone array to generate stimuli for the subjective study and for analysis of the sound fields. Seven halls of various sizes, shapes, and uses were measured to ensure a wide range of LEV conditions for this study.

4.2.1 Measurement Hardware

The hardware used to measure room IRs included an mh acoustics em32 Eigenmike spherical microphone array and a three-way omnidirectional sound source. A Crown XLS 2500 audio amplifier was used to drive a three-way sound source and an RME Babyface was used as a D/A converter to send audio signals to the amplifier.

4.2.1.1 Spherical microphone array

The em32 Eigenmike contains 32 1.3-cm (0.5-inch) diameter omnidirectional microphones mounted on a rigid sphere with an 8.4-cm (3.3-inch) diameter. The Eigenmike array is capable of measuring spherical harmonics up to 3rd order. Along with the microphone array, the Eigenmike system includes the Eigenmike Interface Box (EMIB), which interfaces with a laptop via firewire and an ASIO driver. The EMIB allows for measurements with a synchronous input and output by using the ADAT output. For this study, the ADAT output was sent to the RME Babyface in order to output an analog signal for the amplifier.

4.2.1.2 Three-way omnidirectional sound source

A three-way omnidirectional sound source was used to excite each hall for IR measurements, shown in Figure 4-2. The sound source consists of a low-frequency subwoofer containing two 25-cm (10-inch) drivers in a sealed box (40 - 120 Hz), a mid-frequency dodecahedron with 12 10-
cm (4-inch) mid-bass drivers (120 Hz – 1.5 kHz), and a high-frequency dodecahedron made with 12 closely spaced 1.9-cm (0.75-inch) dome tweeters (1.5 – 20 kHz). The three-way source is omnidirectional up to 5 kHz. Free-field measurements were obtained of the subwoofer both in a large open area outdoors, a concrete-paved parking lot, using the ground-plane measurement technique [82], and in an anechoic chamber. These measurements were used to generate minimum-phase equalization filters [83] in order to have a flat frequency response (averaged over angle above 5 kHz).

For each source-receiver pair, three separate IRs were measured for each of the three sound sources one-at-a-time in order to position the centers of each loudspeaker at the same \((x,y,z)\) source measurement position. Colocation of the loudspeakers was deemed necessary after a series of measurements in a 2500-seat auditorium, which showed that when the three loudspeaker components were stacked, differences in the radiation pattern of the sources arose due to the scattering off of the other components. A three-way crossover was then applied to the three IRs comprised of linear phase FIR (finite impulse response) filters which roll off at 48 dB/octave at 120 Hz and 1.5 kHz, respectively. The crossover filters were designed such that they sum to a delayed impulse in the time domain and have a flat magnitude response in the frequency domain.

### 4.2.2 Details of the Seven Measured Halls and the Receiver Locations

Room IRs were obtained in seven different halls of various sizes and shapes. Details of each of the halls are summarized in Table 4-1, and approximate dimensions of the hall are summarized in Table 4-2. Hall 1, the 400 seat shoebox, featured variable acoustics in the form of retractable
curtains – condition (a) was the most absorptive configuration and condition (b) was the most reverberant configuration. IRs were obtained in 12-20 receiver locations in each hall, though only one receiver location from each hall was used in the subjective study. Measurements were obtained using consistent source-receiver distances from 10 to 25 m (33 to 82 ft) in 5 m (16.4 ft) steps, as hall size allowed, as shown in Figure 4-3. For each source-receiver distance, measurements were obtained at 0 m, 5 m, and 10 m from the center of the hall along the width of the dimension. Each sound source was placed in the middle of each stage at a height of 1.5 m and a distance of approximately 2.5 m from the front of the stage.

Table 4-1: Details about the seven halls measured as a part of this study.

<table>
<thead>
<tr>
<th>Hall Number</th>
<th>Label</th>
<th>Shape</th>
<th>Primary Use</th>
<th>Seats</th>
<th>Midband T30(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a¹</td>
<td>Shoe 400a</td>
<td>Shoebox</td>
<td>Recital Hall</td>
<td>400</td>
<td>1.2</td>
</tr>
<tr>
<td>1b¹</td>
<td>Shoe 400b</td>
<td></td>
<td></td>
<td></td>
<td>1.8</td>
</tr>
<tr>
<td>2</td>
<td>Shoe 571</td>
<td>Shoebox</td>
<td>Multi</td>
<td>571</td>
<td>0.8</td>
</tr>
<tr>
<td>3</td>
<td>Shoe 600</td>
<td>Shoebox</td>
<td>Multi</td>
<td>600</td>
<td>1.2</td>
</tr>
<tr>
<td>4</td>
<td>Fan 726</td>
<td>Fan</td>
<td>Lecture Hall</td>
<td>726</td>
<td>1.2</td>
</tr>
<tr>
<td>5</td>
<td>Shoe 1200</td>
<td>Shoebox</td>
<td>Concert Hall</td>
<td>1200</td>
<td>1.8</td>
</tr>
<tr>
<td>6</td>
<td>Shoe 1302</td>
<td>Shoebox</td>
<td>Multi</td>
<td>1302</td>
<td>1.1</td>
</tr>
<tr>
<td>7</td>
<td>Fan 2100</td>
<td>Fan</td>
<td>Multi</td>
<td>2100</td>
<td>1.4</td>
</tr>
</tbody>
</table>

Table 4-2: Approximate dimensions the seven halls measured as a part of this study.

<table>
<thead>
<tr>
<th>Hall Number</th>
<th>Label</th>
<th>Length (m) / (ft)</th>
<th>Width (m) / (ft)</th>
<th>Height (m) / (ft)</th>
<th>Volume (m³) / (ft³)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a¹</td>
<td>Shoe 400a</td>
<td>18 / 60</td>
<td>16 / 54</td>
<td>10 / 34</td>
<td>3,000 / 110,000</td>
</tr>
<tr>
<td>1b¹</td>
<td>Shoe 400b</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Shoe 571</td>
<td>20 / 65</td>
<td>16 / 54</td>
<td>9 / 31</td>
<td>3,000 / 110,000</td>
</tr>
<tr>
<td>3</td>
<td>Shoe 600</td>
<td>16 / 53</td>
<td>27 / 91</td>
<td>9 / 29</td>
<td>4,000 / 140,000</td>
</tr>
<tr>
<td>4</td>
<td>Fan 726</td>
<td>24 / 79</td>
<td>26 / 85</td>
<td>10 / 33</td>
<td>6,300 / 220,000</td>
</tr>
<tr>
<td>5</td>
<td>Shoe 1200</td>
<td>32 / 105</td>
<td>21 / 70</td>
<td>17 / 57</td>
<td>11,800 / 420,000</td>
</tr>
<tr>
<td>6</td>
<td>Shoe 1302</td>
<td>34 / 111</td>
<td>22 / 71</td>
<td>13 / 42</td>
<td>9,300 / 330,000</td>
</tr>
<tr>
<td>7</td>
<td>Fan 2100</td>
<td>31 / 103</td>
<td>29 / 96</td>
<td>15 / 49</td>
<td>13,600 / 480,000</td>
</tr>
</tbody>
</table>

¹ The 400 seat Shoebox was measured in two absorption conditions, where (a) was the most absorptive condition and (b) was the most reverberant condition.
² Midband T30 is the average reverberation time in the 500 Hz to 2 kHz octave bands.
4.3 SUBJECTIVE STUDY USING AMBISONIC REPRODUCTIONS OF SPATIAL IRS

The purpose of the subjective study was to evaluate the LEV of the measured room IRs. The room IRs obtained in the seven halls were used in the subjective listening test in which participants listened to stimuli and rated how enveloped they felt by the sound field. The subjective test was organized into four test sets. The purpose of the first two sets was to evaluate the effect of overall level on LEV. Set 1 consisted of stimuli generated from the room IRs as measured, including level differences, while. Set 2 was generated using the same IRs equalized for overall A-weighted level. Sets 3 and 4 were generated using hybrid IRs containing portions of enveloping and unenveloping IRs to study the dependence on the arrival time of directional energy with respect to LEV.

The stimuli were presented to subjects in a paired comparison style test using a graphical user interface (GUI) implemented in Max 7 [75]. A pair of stimuli was given to each subject with the order of A and B randomized. For each pair, the participants then instructed to chose whether
they felt sound field A or B was more enveloping. After making their selection, the subjects were then asked on a continuous rating scale from 0 to 10 how much more enveloping their selection was than the other stimulus. The anchor points on the rating scale were 0: “A and B are both equally enveloping”, 2.5: “A/B is somewhat more enveloping than B/A,” 5: “A/B is more enveloping than B/A,” 7.5: “A/B is much more enveloping than B/A,” and 10: “A/B is extremely more enveloping than B/A.” This style of test was chosen because in a previous study comparing different listening test methods including several paired comparison variants, results suggested that the paired comparison method with a continuous rating scale yielded the most consistent results, i.e. the standard deviation of the ratings was the lowest for this type of test [84]. The main disadvantage of this testing style is that the testing time is increased over other methods.

The entire testing session lasted roughly one hour per subject. At the beginning of the test, participants completed a short training period. This training period contained a tutorial explaining the GUI, with explicit instructions to focus only on LEV while ignoring other aspects of the sound field, including ASW. LEV was defined as the sense of feeling surrounded by the sound or immersed in the sound. Following the tutorial, the participants were given a training set to familiarize themselves with the GUI. The subjects were then instructed that the test was beginning and were given a full set of 28 A/B pairs, however this set served as additional practice. The data from this practice set were not used in the analysis. After the practice set, the subjects were presented with four test sets, described in Section 4.3.1, presented in a randomized order. The participants were given two 5-minute breaks after approximately 20 minutes of testing in between sets.

The subjective test was conducted with 21 participants. Of the 21 participants, 2 were removed as outliers due to inconsistent ratings resulting in a total of 19 subjects (11 male, 8 female). Participants were required to have hearing thresholds of 15 dBHL from 250 Hz to 8 kHz. Subjects were also required to be musicians with at least 5 years of formal music training and be an active musician (e.g. performing in an ensemble or taking music lessons). The average age of the participants was 25 years old, and subjects had an average of 9 years of formal music training.

4.3.1 Subjective listening test stimuli

The subjective listening test stimuli were generated from the IRs described in Section 4.2.2. These IRs were processed for 3rd order Ambisonics playback over the 30-loudspeaker array in an
anechoic chamber. The IRs were rotated such that the direct sound was directly in front of the listener at 0° azimuth and 90° elevation (i.e. the horizontal plane). Informal listening tests of the Ambisonic IRs suggested that the LEV differences in different receiver positions within a single hall were much smaller than the LEV differences between halls. Therefore, one IR was used from each hall. The IRs used from each hall had the same source-receiver distance of 15 m and were measured in the center of the hall along the width as shown in Figure 4-3. This receiver location was selected because informal listening tests showed that this distance struck a good balance between the direct sound field and the reverberant sound field.

The IRs were convolved with a motif from the anechoic recording of Mozart’s Don Giovani (an aria of Donna Elvira) [85]. All instruments for this motif were recorded separately, however, only the string instrument parts were used in this study (violins 1 and 2, cello and bass). Furthermore, the IR convolved with each instrument was modified such that the direct sound of each instrument was spatially separated. This modification was done by rotating the direct sound in azimuth from -4° to 4° in 2° steps. Only the direct sound was modified. This process was done in order to create a more realistic auralization since all IR measurements had been obtained using an omnidirectional source placed in the center of the stage. The auralized IRs were presented in four test sets, described below, which were presented to each subject in a randomized order.

4.3.1.1 Test set 1: IRs from Seven different halls as measured
Set 1 consisted of eight stimuli generated from the seven different halls including the two absorption settings for Hall 1. The objective of this test set was to obtain baseline LEV ratings of the halls without manipulating the IRs. Each IR utilized in this set was measured using a source-receiver distance of 15 m and was measured in the center of the hall along the width. The stimuli were generated using the IRs as measured without modification, maintaining the absolute level differences between each IR. A/B pairs were created in which every stimulus was compared to every other stimulus, resulting in 28 pairs.

4.3.1.2 Test set 2: IRs from Seven different halls equalized for level
Set 2 was generated using the same eight stimuli from Set 1, however a scale factor was applied to each IR to equalize the overall A-weighted levels of all stimuli. The purpose of this test set was to study the effect that overall level has on LEV ratings without altering the spatial characteristics of the IRs. The scale factors were calculated by convolving the omnidirectional
IRs with pink noise, applying an A-weighting filter, and calculating an RMS level. The IRs were then divided by the RMS level. The range of the level adjustments was ±3 dB. This test set also contained 28 pairs.

4.3.1.3 Test set 3: High-LEV (HLEV) and Low-LEV (LLEV) hybrid IRs

Set 3 contained stimuli created with hybrid IRs, which contained portions of two different room IRs. Informal listening tests identified one IR from Set 2 that was judged to be unenveloping (Hall 7), and one IR from Set 2 that was judged to be highly enveloping (Hall 6). These hybrid IRs will be referred to as “High-LEV” (HLEV) and “Low-LEV” (LLEV), respectively. For this test set, hybrid IRs were created that contained the early part of the HLEV IR, and the late part of the LLEV IR (referred to as “HLEV&LLEV hybrid IRs”) as shown in Figure 4-4. The goal of this test set was to evaluate the dependence on crossover time when transitioning from HLEV to LLEV, informing integration limits for the LEV metric. The two IRs were crossed-over in time using 2.5 ms half Hann windows. Five IRs were generated using early-late crossover times of 40 ms, 60 ms, 80 ms, 120 ms, and 200 ms. In addition to the five HLEV&LLEV hybrid IRs, the entire individual HLEV and LLEV IRs were included in this test set, resulting in a total of 7 stimuli (21 pairs).
4.3.1.4 Test set 4: LLEV&HLEV hybrid IRs

Set 4 contains hybrid IRs similar to Set 3, although in this set, the hybrid IRs contain the early part of the LLEV IR, and the late part of the HLEV IR (referred to as “LLEV&HLEV hybrid IRs”). The purpose of this test set was to again evaluate crossover time dependence to inform integration limits for the LEV metric. This set contains five LLEV&HLEV hybrid IRs using early-late crossover times of 40 ms, 60 ms, 80 ms, 120 ms, and 200 ms, similar to Set 3, in addition to the complete HLEV and complete LLEV IRs. This set also contains 7 stimuli and a total of 21 pairs.

4.4 SUBJECTIVE STUDY RESULTS

4.4.1 Statistical Analysis of LEV Ratings

A statistical analysis was performed to convert the paired comparison scores to overall LEV ratings for each test set. The overall LEV ratings were then used to evaluate any statistically significant differences between the average LEV ratings within each test set. The overall LEV rating scores were calculated for each individual subject by summing pairs of ratings [84]:

![Figure 4-4: High-LEV (HLEV) and Low-LEV (LLEV) hybrid IR with a crossover time of 200 ms.](image-url)
\[ \text{Rating}_i = \sum_j p_{ij} S_{ij}. \]  

(4-5)

where \( p_{ij} \) is the probability that stimulus \( i \) is more enveloping than stimulus \( j \) (either 0 or 1 in this case since there is a single observation for a single subject), and \( S_{ij} \) is the magnitude of the LEV rating between stimulus \( i \) and stimulus \( j \). As an example, if stimulus A was judged to be more enveloping than stimulus B and assigned a score of 5, then the contribution of that pair to the overall rating of A would be 5 \((1 \times 5)\), and the contribution to the overall rating of B would be 0.

The method of summing pairs of ratings was chosen as it has the advantage over the traditional paired-comparison analysis techniques of Thurstone’s model V or the Bradley-TL model in that it yields rating scores for each individual subject rather than overall ratings [86]. These data can then be used for more advanced statistical analysis techniques [84] including analysis of variance (ANOVA) [65]. For this particular study, a one-way repeated measures ANOVA was performed on the obtained subjects’ ratings for each of the test sets, along with pairwise t-tests with Tukey corrections for each pair within a set [65]. The resulting ratings were normalized on a scale from 0 to 1, where a rating of 0 is the minimum envelopment, i.e. not more enveloping than any other stimulus, and a rating of 1 is the maximum envelopment, i.e. extremely more enveloping than every other stimulus.

4.4.2 Results: Set 1 – IRs from 7 different halls as measured

Set 1 had a large range of LEV ratings, ranging from 0.04 to 0.73 on a scale from 0 to 1 as shown in Figure 4-5 where a rating of 0 is the minimum envelopment, i.e. not more enveloping than any other stimulus, and a rating of 1 is the maximum envelopment, i.e. extremely more enveloping than every other stimulus. The highest average LEV rating of 0.73 ± 0.03 was found for Hall 1b, the 400-seat recital hall in the reverberant setting. The lowest average LEV rating of approximately 0 was found for three halls: Hall 6, the 1302 seat shoebox hall (0.04 ± 0.01), Hall 2, the 571 seat shoebox hall (0.06 ± 0.01), and Hall 7, the 2100 seat fan shaped hall (0.07 ± 0.01). The highlighted colored-shape indicates statistically significant differences in pairs of ratings at \( p < 0.05 \), where differences between ratings that share the same highlighted colored-shape are not statistically significant. For example, Halls 1a, 3, and 5 do not have statistically significant differences in mean LEV rating, and thus are each tagged with a pink triangle highlight. Halls 1a
and 2 have a statistically significant difference in mean LEV rating, and do not share a highlighted color or shape.

Figure 4-5: LEV ratings for Set 1: IRs from 7 different halls as measured. Colored-shapes were added to indicate statistically significant pairs at $p < 0.05$, where stimuli that share the same colored-shape are not significantly different (note that some data points have multiple colored-shapes).

A regression analysis was conducted in which late lateral energy level ($L_l$) was used as a predictor of LEV ratings, shown in Figure 4-6. In Figure 4-6 and all subsequent plots, $L_l$ is shown relative to the mean $L_l$ in each test set in order to highlight the range of the measured values. A high correlation was found with a correlation coefficient of $R^2 = 0.94$, $p < 0.001$. However, further analysis showed that the correlation was primarily due to the level differences in each of the halls. Using overall A-weighted level as a predictor, the correlation coefficient was found to be $R^2 = 0.90$, $p < 0.001$. Moreover, using late A-weighted level as a predictor, the correlation
coefficient was found to be $R^2 = 0.96$, $p < 0.001$, indicating that in a wide range of halls, late level alone is a good predictor for LEV.

![Figure 4-6: Set 1 LEV ratings vs. late lateral energy level. LEV ratings were found to have a high correlation with $L_j$, although this correlation is primarily due to level differences between the stimuli.](image)

4.4.3 Results: Set 2 – IRs from 7 different halls equalized for level

The results of Set 2, shown in Figure 4-7, give further confirmation that LEV ratings are highly dependent on level since this test set was normalized for overall A-weighted level. Comparing the results from Set 2 to the results of Set 1, the range of LEV ratings has decreased from about [0, 0.73] to a range of [0, 0.53], indicating that the LEV differences in this set were smaller than the LEV differences in Set 1. Additionally, after normalizing for overall A-weighted level, the LEV ratings changed significantly. Hall 5, the 1200 seat shoebox hall, has the highest LEV rating in this test set (0.53 ± 0.04), whereas in Set 1, this hall was in the middle of the range of LEV ratings (0.36 ± 0.03). Similarly, Hall 1b, the 400-seat shoebox hall in the most reverberant setting, which had the highest LEV rating in Set 1 (0.73 ± 0.03) is in the middle of the LEV range in Set 2 (0.32 ± 0.02). These results further indicate that overall level is a large contributing factor to the perception of LEV.
Figure 4-7: LEV ratings for Set 2: IRs from 7 different halls, which are the same halls as used in Set 1, but equalized for level. Colored-shapes were added to indicate statistically significant pairs at $p < 0.05$, where stimuli that share the same colored-shape are not significantly different (note that some data points have multiple colored-shapes).

A regression analysis was also conducted for the Set 2 results using $L_I$ as a predictor for LEV ratings, shown in Figure 4-8. In this test set, equalizing the A-weighted level resulted in the correlation coefficient decreasing to $R^2 = 0.77$, $p < 0.004$ from $R^2 = 0.94$, $p < 0.001$ in Set 1. This result indicates that when level differences are small, $L_I$ does not perform as well in predicting LEV differences.
Figure 4-8: Set 2 LEV rating vs. $L_i$. The correlation between $L_i$ and LEV rating for Set 2 ($R^2 = 0.77$, $p < 0.004$) is lower than in Set 1 ($R^2 = 0.94$, $p < 0.001$) after normalizing the overall A-weighted level.

4.4.4 Results: Set 3 – H-LEV&L-LEV Hybrid IRs

As previously mentioned, the goal of Set 3 was to evaluate the time dependence of LEV by combining the early part of an enveloping IR (HLEV) with the late part of an unenveloping IR (LLEV), and also varying the crossover time. Set 3 results, seen in Figure 4-9, show that in the H-LEV&L-LEV hybrid test set, as the crossover time increases, the LEV ratings increase. For crossover times of 80 ms and lower, the LEV ratings are not significantly different than the LEV ratings of the whole L-LEV IR. Increasing the crossover time from 80 ms to 120 ms increases the average LEV rating from $0.14 \pm 0.01$ to $0.25 \pm 0.03$ and further increasing the crossover time to 200 ms increases the LEV rating to $0.33 \pm 0.03$. The LEV rating at 200 ms is approximately 57% of the LEV rating for the whole H-LEV IR.

A regression analysis was also run using $L_i$ as a predictor for LEV ratings, shown in Figure 4-10. Since the average LEV ratings were only found to increase at crossover times greater than 80 ms, which is the same crossover time as $L_i$, the results of this test set had a high correlation where $R^2 = 0.95$, $p < 0.001$. 
Figure 4-9: LEV ratings for Set 3. Colored-shapes were added to indicate statistically significant pairs at $p < 0.05$, where stimuli that share the same colored-shape are not significantly different. LEV ratings for crossover times 80 ms and lower are not significantly different than the LEV rating for the whole L-LEV IR.
4.4.5 Results: Set 4 – L-LEV\&H-LEV Hybrid IRs

As described above, the goal of Set 4 was to evaluate the time dependence of LEV by combining the early part of an unenveloping IR (LLEV) with the late part of an enveloping IR (HLEV) and varying the crossover time. The Set 4 results from the L-LEV\&H-LEV hybrid IRs show a very different trend than the Set 3 results, as seen in Figure 4-11. Although the Set 3 results show that LEV doesn’t increase until beyond the 80 ms crossover time, the Set 4 results show that manipulating the early sound field has a significant effect on the perceived envelopment.

Replacing the first 40 ms of the H-LEV IR with the L-LEV IR drops the LEV rating from $0.42 \pm 0.03$ to $0.32 \pm 0.02$. Additionally, replacing the first 80 ms of the H-LEV IR with the L-LEV IR reduces the LEV rating from $0.42 \pm 0.03$ to $0.19 \pm 0.02$, which is less than half of the original H-LEV IR LEV rating. Although the results from Sets 3 and 4 are seemingly contradictory, these results are consistent with the findings of a previous study by the Dick and Vigeant [87] using the hybrid IR technique (see Chapter 3 Section 3.6).
Figure 4-11: LEV ratings for Set 4. Colored-shapes were added to indicate statistically significant pairs at p < 0.05, where stimuli that share the same colored-shape are not significantly different.

As a result of the large effect that manipulating the early sound field has on LEV ratings, the correlation between LEV ratings and $L_i$ is much lower than in Set 3 with a $R^2 = 0.73$, p = 0.015. Crossover times ranging from 0 to 80 ms all have the same value of relative $L_i$ of +1.0 dB, which manifests itself as a vertical line in the plot of LEV ratings vs. $L_i$, as shown in Figure 4-12. Results from this test set further indicate that $L_i$ is not adequate in predicting LEV in certain situations.
Figure 4-12: Set 4 LEV rating vs. $L_J$. Each data point is labeled for its crossover time. Correlation is much lower than in previous sets ($R^2 = 0.73$) because the four data points up to the 80 ms crossover point have identical values of $L_J$ but very different LEV ratings.
4.5 METRIC DEVELOPMENT

The results from the four subjective test sets were used to inform metric development. The metric was developed using energy integrated as a function of azimuthal angle, elevation angle, and time for octave bands from 125 Hz to 4 kHz. The goal was to find the integration limits and frequency bands that maximize the correlation between integrated energy and overall LEV ratings for the four test sets. The findings from this subjective study results along with those from previous studies indicate that the metric should have a strong level component and include a portion of the early sound field.

The first step was to compute the directional IRs utilizing a 3rd order plane wave decomposition (PWD) on the spherical microphone array IRs to yield directional IRs. Directional IRs were generated for each of the IRs over the entire sphere using look directions every 3 degrees in azimuth and elevation, yielding a grid of 7,200 directional IRs. These IRs were then time windowed using a rectangular window in 5 ms segments, and subsequently filtered into octave bands from 125 Hz to 4 kHz. The energy in each time window was summed, resulting in one energy level per grid point in each time segment for each octave band. This processing yields energy which can be summed as a function of time, azimuth angle, elevation angle, and frequency. An example grid is shown in Figure 4-13 for the 1200 seat shoebox hall (Hall 5). This grid is the energy in the 1 kHz octave band summed from 60 ms to 200 ms in dB relative to the maximum.

![Figure 4-13: Energy grid for Hall 5, the 1200-seat shoebox hall, in the 1 kHz octave band summed from 60 ms to 200 ms. Energy is shown in dB relative to the maximum. With 0° azimuth pointing toward the stage and 0° elevation pointing straight up. The images on the left show the energy distributions over spheres, while the image on the right shows the same information, but flattened onto a 2-D plot (similar to an unraveled map.).]
To develop the metric, the energy was integrated iteratively over azimuth, elevation, and time using different integration limits for each dependence. The integration was approximated using the trapz function in MATLAB. The effects of each of these three variables on the integration limits are discussed in the sections below. In order to optimize the integration limits, the parameters of azimuth, elevation, and time were varied in order to maximize the correlation between LEV rating and the integrated energy:

\[
\int_{\phi_1}^{\phi_2} \int_{\theta_1}^{\theta_2} \int_{t_1}^{t_2} p_{grid}(t, \phi, \theta) \sin(\theta) \, dt \, d\theta \, d\phi,
\]

where \(\phi_1\) and \(\phi_2\) are the azimuthal integration limits, \(\theta_1\) and \(\theta_2\) are the elevation integration limits, and \(t_1\) and \(t_2\) are the time integration limits.

In the following sections, the metric development is shown with data from Set 2, which is the set that had the largest impact on changing integration limits. The data from Set 2 were most sensitive to changes since the stimuli in this set were level equalized, which yielded smaller LEV differences than Set 1, and because Set 2 contains stimuli from all of the halls. This iterative process was conducted for the 125 Hz to 4 kHz octave bands.

### 4.5.1 Azimuthal Angular Dependence

The azimuthal angular dependence was investigated by integrating energy over the sphere while changing the azimuthal integration limits. The elevation and time integration limits were kept constant at 0 and \(\pi\), and 60 and 200 ms, respectively. The 200 ms time cutoff was chosen as a starting point because the sound field after 200 ms in general has less variation in energy as a function of azimuth angle. As an example, Shoe 1200 (Hall 5) has roughly an 8 dB range in energy integrated from 60 ms to 200 ms (see Figure 4-13), whereas the energy integrated from 200 ms to 500 ms has roughly a 4 dB range (see Figure 4-14). The azimuthal angular dependence was investigated in two separate steps: first by rejecting progressively more of the front energy, and second by rejecting progressively more of the rear energy.
4.5.1.1 LEV Rating Correlation Excluding Frontal energy

The energy over the sphere was integrated by varying the azimuthal integration limits rejecting the front energy symmetrically:

\[
\int_{\alpha}^{360^\circ - \alpha} \int_{0^\circ}^{180^\circ} \int_{60\text{ ms}}^{200\text{ ms}} p_{\text{grid}}^2(t, \phi, \theta) \sin(\theta) \, dt \, d\theta \, d\phi,
\]

where the angular limit \(\alpha\) was varied from 0 to 90 degrees in 5\(^\circ\) steps. For each of these 5\(^\circ\) steps, a regression analysis was performed in which the LEV rating was predicted by the integrated energy. As shown in Figure 4-15, the correlation coefficient was maximized in this case at \(\alpha = 20^\circ\) (indicated by the circle in in Figure 4-15), where \(R^2 = 0.77\) and \(p < 0.005\). In other words, the optimal azimuthal angle range for the metric excludes energy from -20 to +20 degrees. In Figure 4-15 and subsequent figures, the 1 kHz octave band results are shown as it was found to have the highest correlation with LEV.

4.5.1.2 LEV Rating Correlation Excluding Rear energy

Similarly, energy was integrated over time, azimuth, and elevation while iteratively rejecting the rear energy symmetrically:

\[
\left( \int_{0^\circ}^{180^\circ - \alpha} + \int_{180^\circ + \alpha}^{360^\circ} \right) \left( \int_{0^\circ}^{180^\circ} \int_{60\text{ ms}}^{200\text{ ms}} p_{\text{grid}}^2(t, \phi, \theta) \sin(\theta) \, dt \, d\theta \right) d\phi,
\]

\[ (4-8) \]
where the angular limit $\alpha$ was varied from 0° to 90° degrees. For each of these 5° steps, a regression analysis was performed in which the LEV rating was predicted by the integrated energy. The correlation coefficient was maximized in this case at $\alpha = 70°$ (indicated by the circle in Figure 4-15). In other words, the optimal azimuthal angle range for the metric excludes energy from 110° to 250°. As shown in Figure 4-15, the correlation coefficient is maximized when a much larger angular section is removed from the rear than the front. Additionally, the correlation coefficient is much higher rejecting the rear energy ($R^2 = 0.89$, $p < 0.001$) than the front energy ($R^2 = 0.77$, $p < 0.005$), indicating that rejecting rear energy is more important than rejecting front energy to predict LEV. This finding agrees with early LEV experiments [67] [10] but is contradictory to Ref. [13], which states that rear energy is an important contributor to LEV.
Figure 4-15: Correlation coefficient vs. rejection angle for Set 2 in the 1 kHz octave band for both front sound rejection (blue line) and rear sound rejection (orange line). The maximum $R^2$ values are circled for each case. The correlation for rejecting the front sound is maximized at 20°. The correlation for rejecting the rear sound is maximized at 70 degrees, which is a larger rejection angle than was found for the front energy. Additionally, rejecting the rear sound has a higher correlation coefficient than rejecting the front sound.

4.5.2 Elevation Angular Dependence

The elevation angular dependence was investigated by integrating energy over the sphere while changing the elevation integration limits and keeping the azimuthal limits and time limits constant,

$$
\left( \int_{20^\circ}^{120^\circ} + \int_{240^\circ}^{340^\circ} \right) \left\{ \int_{\theta_1}^{\theta_2} \int_{60\,\text{ms}}^{200\,\text{ms}} p_{\text{grid}}^2(t, \phi, \theta) \sin(\theta) \, dt \, d\theta \right\} d\phi,
$$

where the azimuthal limits applied were those from the first phase of this analysis. The investigation of elevation angle was conducted in two steps similar to the investigation of azimuthal angle. First, $\theta_1$ was held constant at 0° while $\theta_2$ was varied from 90° to 180° in 5° increments. Next, $\theta_2$ was held constant at 180° while was varied from 0° to 90° in 5° increments.
The correlation coefficient was maximized for $\theta_1 = 30^\circ$ and $\theta_2 = 130^\circ$. However, optimizing the elevation limits had a smaller impact on the correlation between integrated energy and LEV ratings than optimizing the azimuthal limits. Optimizing the elevation limits in the 1 kHz octave band yields $R^2 = 0.93$ ($p < 0.001$), whereas integrating the full elevation span yields $R^2 = 0.92$ ($p < 0.001$).

### 4.5.3 Time Dependence

The time dependence was investigated by keeping the angular integration limits constant, using the values found in the azimuthal dependence correlation analysis, and varying the early time cutoff and the late time cutoff in order to maximize the correlation coefficient between the metric and the LEV ratings.

#### 4.5.3.1 LEV Rating Correlation Varying Early Time Cutoff

The optimal early cutoff time was analyzed by integrating the energy from $t_1$ through the end of the IR:

$$
\left( \int_{20^\circ}^{120^\circ} + \int_{240^\circ}^{340^\circ} \right) \left\{ \int_{0^\circ}^{180^\circ} \int_{t_1}^{\infty} p^2_{grid}(t, \phi, \theta) \sin(\theta) \ dt \ d\theta \right\} d\phi
$$

(4-10)

where $t_1$ was varied from 0 ms to 100 ms in increments of 5 ms. The correlation coefficient between the integrated energy and LEV rating is shown in Figure 4-16 for the 1 kHz and 2 kHz octave bands. The correlation coefficient was maximized from 30 to 70 ms, depending on frequency, where $R^2_{\text{max @ 1kHz}} = 0.96$ ($p < 0.001$) at 30 ms and $R^2_{\text{max @ 2kHz}} = 0.80$ ($p < 0.005$) at 70 ms. For the purposes of this study, the 60 ms cutoff time was used in the final metric as a compromise between the 1 and 2 kHz octave bands to maximize correlation since the correlation coefficient at 60 ms in the 1 kHz octave band is only slightly lower at 60 ms than at 30 ms. Additionally, 60 ms was chosen as a compromise between the contradictory results of Set 3, which would imply a cutoff time of 80 ms (see Figure 4-9), and Set 4, which would imply a cutoff time of 40 ms or earlier (see Figure 4-11). Further studies may be required to optimize the early integration time limit and it is possible that a frequency-dependent time cutoff is necessary.
4.5.3.2 LEV Rating Correlation Varying Late Time Cutoff

The late cutoff time was optimized by keeping $t_1$ constant at 60 ms, and varying the late time cutoff, $t_2$, from 150 ms to 500 ms in steps of 10 ms:

$$
\left\{ \int_{180^\circ}^{340^\circ} \int_{120^\circ}^{340^\circ} \int_{0^\circ}^{180^\circ} \int_{0^\circ}^{t_2} p_{grid}(t, \phi, \theta) \sin(\theta) \, dt \, d\theta \right\} d\phi
$$

The cutoff time of 150 ms was chosen as the starting time because the results of Sets 3 and 4 indicate that there is information important to the perception of LEV beyond 200 ms. The correlation coefficient steadily increases as more of the IR is integrated and asymptotically approaches a maximum at 400 ms, as shown in Figure 4-17. Therefore, 400 ms was selected as the late cutoff time for this metric. However, for the rooms considered in this study, increasing the late cutoff time did not have an effect on the integrated energy, so these results don’t contradict the commonly used late cutoff time of the end of the IR. An increased cutoff time
may be beneficial for rooms with a longer reverberation time than the halls measured in this study.

Figure 4-17: Correlation coefficient vs. late time cutoff. Correlation increases slightly as the late time cutoff is increased, and asymptotically reaches a maximum around 400 ms.

4.5.4 Overall Metric

Using the optimized integration limits from Sections 4.5.1 and 4.5.3, the proposed new metric for listener envelopment, which is both dependent on time and arrival directions, is the mid-late spatial energy metric, $J_S$:

$$J_S = \left( \int_{120^\circ}^{20^\circ} + \int_{340^\circ}^{240^\circ} \right) \left( \int_{130^\circ}^{340^\circ} \int_{400 \text{ ms}}^{120 \text{ ms}} \int_{30^\circ}^{60^\circ} p_{grid}^2(t, \phi, \theta) \sin(\theta) \, dt \, d\theta \right) \, d\phi. \quad (4-12)$$

The proposed notation, $J_S$, is to maintain consistency with the spatial metrics in Appendix A of the ISO 3382 room acoustics measurement standard, which all contain the letter $J$ [43] and the $S$ notation is short for mid-late spatial energy.
$J_S$ was calculated for all four test sets, per octave band, and for $1^{st}$, $2^{nd}$, and $3^{rd}$ order PWDs. Due to limitations of the spherical microphone array at low frequencies, $2^{nd}$ order PWDs were used only in the 500 Hz to 4 kHz octave bands, and $3^{rd}$ order PWDs were used only in the 1 kHz to 4 kHz octave bands. The correlation coefficients for the four test sets in each octave band are presented in Table 4-3. All regressions were significant at $p < 0.05$, except in Set 2 where non-significant regressions are labeled NS. The highest correlations were found in the 1 kHz and 2 kHz octave bands for all of the sets.

Table 4-3: Correlation coefficients between the new metric and mean LEV of the four test sets. All correlation coefficients shown are significant at $p < 0.05$ and the non-significant regressions are denoted by “NS”.

<table>
<thead>
<tr>
<th>Set 1: $R^2$ Values for Each Octave Band [Hz]</th>
<th>Set 2: $R^2$ Values for Each Octave Band [Hz]</th>
<th>Set 3: $R^2$ Values for Each Octave Band [Hz]</th>
<th>Set 4: $R^2$ Values for Each Octave Band [Hz]</th>
</tr>
</thead>
<tbody>
<tr>
<td>125 250 500 1 k 2 k 4k</td>
<td>125 250 500 1 k 2 k 4k</td>
<td>125 250 500 1 k 2 k 4k</td>
<td>125 250 500 1 k 2 k 4k</td>
</tr>
<tr>
<td>1st order SH: 0.87 0.98 0.97 0.98 0.95 0.85</td>
<td>2nd order SH: 0.95 0.97 0.93 0.85 NS NS NS</td>
<td>3rd order SH: 0.98 0.94 0.85 NS NS NS</td>
<td>1st order SH: 0.88 0.92 0.91 0.95 0.92 0.82</td>
</tr>
<tr>
<td>2nd order SH: 0.95 0.97 0.93 0.85</td>
<td>2nd order SH: 0.96 0.97 0.93 0.85 0.87 0.86</td>
<td>3rd order SH: 0.96 0.93 0.85 0.87 0.86 0.86</td>
<td></td>
</tr>
<tr>
<td>3rd order SH: 0.98 0.94 0.85</td>
<td>3rd order SH: 0.96 0.93 0.85</td>
<td></td>
<td>3rd order SH: 0.96 0.93 0.85</td>
</tr>
</tbody>
</table>

One important note is that in Set 2, the level equalized set, high correlations were found primarily in the 1 kHz octave band, and to a lesser extent the 2 kHz octave band. The other octave bands were either not significant or had a low correlation. However in Set 1, which contained large level differences, all octave bands were found to have high correlations with LEV ratings. This finding suggests that the spatial properties of the sound field are well captured by the 1 kHz octave band, which is a region where the auditory system is very sensitive [88].

4.5.4.1 Performance relative to late lateral energy level and strength

As discussed in Section 4.4, the metric Late Lateral Energy Level ($L_J$) performs well in Sets 1 and 3. However, $L_J$ performs worse in Set 2 in which level differences are significantly reduced and in Set 4 in which manipulating the early sound field removes envelopment. The correlation coefficients between the LEV ratings and $L_J$ were computed and are summarized in Table 4-4, along with the correlation coefficients for $J_S$ at the 1 kHz octave band for comparison purposes. As shown, $J_S$ outperforms $L_J$ in all test sets.
In addition to \( L_i \), two additional metrics were computed:

\[
L_{J60} = 10 \log_{10} \left[ \frac{\int_{400\text{ms}}^{60\text{ms}} p_2^2(t)dt}{\int_{0}^{\infty} p_{10}^2(t)dt} \right] \text{[dB]},
\]

which is a modified version of \( L_i \) in which the time integration limits of the dipole term are the optimized limits used in \( J_S \), and:

\[
G_{60} = 10 \log_{10} \left[ \frac{\int_{400\text{ms}}^{60\text{ms}} p^2(t)dt}{\int_{0}^{\infty} p_{10}^2(t)dt} \right] \text{[dB]},
\]

which is a modified version of late strength using the optimized time integration limits and \( p^2(t) \) is the omnidirectional IR. The objective of evaluating \( L_{J60} \) was to determine whether including a portion of the early sound field improves the performance of \( L_i \). The objective of evaluating \( G_{60} \) was to determine if an omnidirectional level term with the proper time integration limits is sufficient for predicting LEV.

The correlation coefficients between the LEV ratings and these metrics were computed and are summarized in Table 4-4. The correlation coefficient of \( L_{J60} \) are slightly improved over \( L_i \) in Set 4 (\( R^2 = 0.81 \) and \( R^2 = 0.72 \), respectively), which suggests that including a portion of the early sound field in \( L_i \) improves the prediction of LEV. However, the correlation coefficients in Sets 2 and 4 are lower than \( J_S \), illustrating the need for improved spatial selectivity. The correlation of \( G_{60} \) is high for Set 1 where the stimuli contained large level differences, but low for Sets 2 through 4 where levels were equalized compared to \( J_S \). This result indicates that large level differences will dominate the perception of LEV over spatial differences, and thus an omnidirectional level term is sufficient as a coarse predictor for LEV in scenarios where large level differences are present.
Table 4-4: Correlation coefficients between LEV rating and $L_j$ and the correlation coefficients between LEV rating and the new metric. All correlation coefficients shown are significant at $p < 0.05$ and the non-significant regressions are denoted by “NS”.

<table>
<thead>
<tr>
<th></th>
<th>Set 1 $R^2$</th>
<th>Set 2 $R^2$</th>
<th>Set 3 $R^2$</th>
<th>Set 4 $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L_j$</td>
<td>0.94</td>
<td>0.77</td>
<td>0.95</td>
<td>0.72</td>
</tr>
<tr>
<td>$J_5$ (1 kHz octave band)</td>
<td>0.97</td>
<td>0.94</td>
<td>0.96</td>
<td>0.87</td>
</tr>
<tr>
<td>$L_{60}$</td>
<td>0.94</td>
<td>0.79</td>
<td>0.93</td>
<td>0.81</td>
</tr>
<tr>
<td>$G_{60}$ (1 kHz octave band)</td>
<td>0.94</td>
<td>0.81</td>
<td>NS</td>
<td>0.66</td>
</tr>
</tbody>
</table>

4.5.4.2 Performance with order reduction

The measurement of 3rd order spherical harmonic components requires a spherical microphone array with a minimum of 16 elements, and commercially available spherical microphone arrays can be cost prohibitive. For this reason, it is desirable to reduce the order to 2nd or 1st order, which require a minimum of 9 elements or 4 elements, respectively. The effect of order reduction was evaluated by generating 2nd and 1st order directional IR grids, and computing the correlation coefficients with LEV ratings, as shown in Table 4-3. The results show that reducing the order of $J_5$ from 3rd order to 2nd or 1st order does not impact the correlation with LEV ratings. In other words, an order-reduced measurement of $J_5$ can be obtained with a 4-element microphone array, and the resulting measurement can still be used to predict LEV with similar performance to the 32-element spherical microphone array.

4.6 Effect of the Early Sound Field

The contradictory results between Sets 3 and 4 (Sections 4.4.4 and 4.4.5) suggest that LEV is affected by the early sound field, but LEV may not be strictly dependent on the amount of early energy. In Set 3, the H-LEV/L-LEV hybrid IR set, the IRs with crossover times up to 80 ms, i.e. including up to 80 ms of the H-LEV IR, are not significantly more enveloping than the using the original L-LEV IR only. Conversely, in Set 4, the L-LEV hybrid IRs exhibit a significant decrease in LEV even when the early time crossovers of 40 ms and 60 ms are used as the transition point for
combining the IRs with low- and high-envelopment ratings. In terms of the early sound field, these results show that by manipulating the IRs, there is a more significant effect on perceived envelopment from modifying early portions of a highly enveloping IR (e.g., the 40 and 60 ms time crossovers in Set 4) than from modifying early portions of an unenveloping IR (e.g., the early time crossovers in Set 3). This asymmetric behavior cannot be strictly accounted for by integrating energy over fixed time and angular windows and, therefore, the LEV metric may be improved by adding an interaction term between the early and late sound fields. As an example, the correlation between LEV and integrated energy can be increased in Set 4 by reducing the early integration time limit to 40 ms ($R^2 = 0.93, p < 0.001$) at 1 kHz, but this occurs at the expense of correlation in Set 3, where $R^2$ is reduced to 0.90, $p < 0.001$, again at 1 kHz.

An additional contributor to the findings in Sets 3 and 4 is that the early reflections of the H-LEV IR are preferable to the early reflections in the L-LEV IR. In the first 40 ms of the IR after the direct sound, the dominant reflections in the H-LEV IR are lateral, whereas for the L-LEV IR, it turns out that the dominant early reflections are coming from the ceiling. Although previous studies have shown that the early sound field affects the perception of apparent source width and not LEV [9], these results show that manipulation of the early reflections do directly impact LEV. This finding suggests that early lateral reflections promote envelopment more so than early ceiling reflections, which is in agreement with previous studies [15].

### 4.7 CONCLUSIONS

The objective of this study was to utilize spatial room IRs to develop a new metric to predict LEV. LEV was investigated using a listening test in which subjects were presented with seven halls and participants rated the LEV of each sound field. The stimuli were presented in four test sets: Set 1, all halls presented as measured; Set 2, all stimuli were equalized for overall A-weighted level; Set 3, hybrid High-LEV & Low-LEV IRs; and Set 4, hybrid Low-LEV & High-LEV IRs. In each set, statistically significant differences in LEV ratings were found. The currently accepted metric of late lateral energy level, $L_i$, was found to be highly correlated with LEV ratings in Sets 1 and 3, but not as well in conditions where differences in level across stimuli were removed (Set 2) and where manipulations of the early sound field affected LEV ratings (Set 4).

The findings in the subjective study were used to inform the development of a new metric to predict LEV based on integration of energy in directional impulse responses. Energy integration
limits in azimuthal angle as well as early and late crossover times were adjusted to maximize the correlation between integrated energy and LEV ratings. The correlation was optimized by rejecting azimuthal directional sound from the front ±20° and from the rear ±70°, and integrating in time from 60 ms to 400 ms. This proposed metric, mid-late spatial energy, $J_S$, outperforms $L_I$ in all conditions tested. The correlation between this metric and LEV remains high even when the metric is calculated using 1st order measurement data only. Accordingly, a 4-microphone (1st-order) array can be used to measure the proposed metric, which are much more widely available and cost effective than 3rd-order microphones.

In halls with a range of sizes and shapes, our results show that most of the differences in LEV can be attributed to level differences within the natural IRs. This finding implies that an omnidirectional level term is sufficient as a coarse predictor for LEV. Further, albeit smaller, differences in LEV exist even after room IRs are level normalized. The differences can be predicted from a more detailed measure operating on the directional aspects of the room IRs. Specifically, this metric includes azimuthal energy from $|20^\circ|$ to $|110^\circ|$, i.e., rejecting a narrow range of the frontal energy and a wide range of the rear energy. The metric also includes a portion of the early energy prior to 80 ms in the IR, specifically from 60 ms onward, which increases the correlation with LEV ratings.

Future work should include an investigation on how early reflections impact the perception of LEV since manipulating the early part of the IRs affected LEV ratings in some cases, e.g. the High-LEV & Low-LEV hybrid IR set, and did not have a significant effect in other cases, e.g. the Low-LEV & High-LEV hybrid IR set. Future work should also include a study on integration limits as a function of frequency.
Chapter 5: Conclusions

5.1 RESULTS SUMMARY

The objective of this work was to utilize room impulse responses (IRs) measured with a 32-
element spherical microphone array to develop a new metric to predict LEV. The first study was
carried out to validate the spherical array measurement setup (Chapter 2 and Ref [37]). The
validation was done by comparing the room acoustics metrics defined in ISO 3382 [43]
measured with a spherical microphone array to measurements obtained with a conventional
microphone configuration, an omnidirectional and figure-8 pair. IR measurements were
obtained in a 2500-seat auditorium using a Brüel & Kjær 4192 12.7 mm (0.5 inch) microphone
and a Sennheiser MKH30 microphone, and an Eigenmike em32 spherical microphone array,
which consists of 32 1.3 cm microphones uniformly distributed over the 8.4-cm diameter
sphere. Six receiver locations were measured in the hall, with three repetitions at each receiver
to evaluate the uncertainty in the positioning of the microphones at each location. The room IRs
were filtered to equalize the frequency response of the microphones and sound source. The
measures evaluated as part of this study were reverberation time \( T_{30} \), early decay time \( EDT \),
clarity index \( C_{80} \), strength \( G \), early lateral energy fraction \( J_{LF} \), and late lateral energy level
\( L_{J} \). The measurements made with both the Eigenmike and the conventional microphone pair
were found to be repeatable for the three repetitions with the exception of EDT in the 4 kHz
octave band at R3 measured with the Eigenmike, which was likely caused by slight
misalignments in re-positioning the spherical array during repetitions.

Differences in the four omnidirectional and two spatial metrics evaluated in this first study
obtained from the IRs between the two microphone configurations were calculated at each
receiver location. The differences in the measures were smaller than 1 JND for most the metrics,
except for C80 at one single receiver position and octave band, and EDT in several receiver
positions at high frequencies (2 kHz and above). These metrics were found to be very sensitive
to small changes in position of the microphones since they are dependent on the early sound
field. The spatial measures \( L_{J} \) and \( J_{LF} \) were found to have smaller differences than EDT and C80
relative to 1 JND. The agreement in the calculated metrics obtained from the IRs of the two
microphone configurations is much better than the agreement reported in previous studies for differing microphone types. The results of the first study show that it is acceptable to measure room acoustic metrics with a spherical microphone in place of the traditional configuration of an omnidirectional and figure-8 microphone pair. This finding gave confidence in the use of spherical microphone arrays for the LEV studies to measure both existing room acoustics metrics and to develop a new metric.

For the first LEV subjective study, room IRs were obtained in the Peter Kiewit Concert Hall in Omaha, NE, USA using the 32-element spherical microphone array, analyzed using beamforming techniques, and processed for 3rd order Ambisonic reproduction to conduct listening tests (Chapter 3, Section 3.4). The Ambisonsics playback system was validated both objectively and perceptually. LEV scores from the subjective listening tests were compared to several objective measures and to the energy in the beamformed directional IRs in different time windows.

To determine whether existing objective metrics were correlated with the perception of LEV, linear regressions were conducted with the LEV ratings as the response variable and individual metrics as the predictors. The predictors found to have the highest correlations were all functions of sound level: strength ($G$), late strength ($G_{\text{late}}$), and late lateral energy level ($L_l$). This finding indicates that throughout this particular hall, the sense of envelopment is highly dependent on level. Metrics having to do with spatial aspects of the room but not overall level, including lateral fraction ($L_f$), late lateral fraction ($L_{LF}$), front-back ratio ($FBR$), and spatially balanced center time ($SBT_S$), were found to have little to no correlation with the LEV ratings.

In comparing the spatial distribution of energy in the IRs to the LEV ratings, it became apparent that in some instances the LEV ratings did not agree with the energy in the late sound field, i.e. pairs of stimuli with statistically different mean LEV ratings, but which had similar late sound fields and vice versa. An exploratory look into different time windows suggested that LEV rating seemed to be correlated with the energy distribution in an earlier time window, specifically from 70 ms to 100 ms.

A second LEV study was conducted using the IRs obtained in the Peter Kiewit Concert Hall to investigate specific time windows in the IRs (Chapter 3, Section 3.6). For this study, hybrid IRs were generated using a mix of an enveloping (HLEV) IR and an unenveloping (LLEV) IR, as well as
hybrid IRs using a mix of the original fully spatial IR (3D) and an IR in which the energy was collapsed to a single direction in front of the listener (mono). For the hybrid IR sets that contained the early part of the HLEV IR and late part of the LLEV IR (abbreviated as “HLEV&LLEV”), as well as the sets that contained the early part of the LLEV IR and the late part of the HLEV IR (LLEV&HLEV), it was found that a crossover time after 120 ms had very little effect on the envelopment. The fact that manipulating the late sound field had little impact on LEV is likely because all of the stimuli came from the same hall, and the late sound field is more consistent throughout the hall than the early sound field. It was also found that a crossover time between 80 and 120 ms in the HLEV&LLEV hybrids, or 60 and 100 ms in the LLEV&HLEV hybrids, resulted in an LEV rating halfway in between the original enveloping and unenveloping IRs. This finding suggests that this time region contains information important to the perception of LEV and that the early part of the impulse response should be taken into account in metrics predicting LEV. Hybrid listening tests containing 3D&mono and mono&3D hybrids confirmed that changes in the early part of the IR impacts the LEV ratings.

A third LEV study was conducted using IRs obtained in seven different halls of various size and shape (Chapter 4). The objective of this study was to use the IRs to develop a new metric to predict LEV. LEV was investigated using a listening test in which subjects were presented with stimuli generated from IRs measured in the seven halls and participants rated the LEV of each sound field. The stimuli were presented in four test sets: Set 1, all halls presented as measured; Set 2, all stimuli were equalized for overall A-weighted level; Set 3, hybrid High-LEV & Low-LEV IRs; and Set 4, hybrid Low-LEV & High-LEV IRs. In each set, statistically significant differences in LEV ratings were found. $L_I$ was found to be highly correlated with the ratings in Sets 1 and 3, but not as well in the other two sets where there were no differences in level across stimuli in Set 2 and manipulating the early sound field affected LEV ratings in Set 4.

The findings in the subjective study were used to inform the development of a new metric to predict LEV, termed mid-late spatial energy, $J_s$: 
\[ J_S = \left( \int_{20^\circ}^{120^\circ} + \int_{240^\circ}^{340^\circ} \right) \left( \int_{30^\circ}^{130^\circ} \int_{60\text{ms}}^{400\text{ms}} p_{grid}^2(t, \phi, \theta) \sin(\theta) \, dt \, d\theta \right) d\phi, \]  

(5-1)

where \( p_{grid}^2 \) is a grid of directional IRs computed via plane wave decomposition (PWD), \( t \) is time, \( \theta \) is elevation angle, and \( \phi \) is azimuthal angle. Energy integration limits in azimuthal angle as well as early and late cutoff times were adjusted until the correlation between LEV ratings and integrated energy were maximized. The correlation was optimized by rejecting the front \( \pm 20^\circ \) and the rear \( \pm 70^\circ \) and integrating in time from 60 ms to 400 ms. This proposed metric outperforms \( L_i \) in all four test sets. The correlation between this metric, \( J_S \), and LEV remains high even when the metric is calculated using 1st order measurement data only. Accordingly, a 4-microphone (1st-order) array can be used to measure the proposed metric; such arrays are much more widely available and substantially more cost effective than 3rd-order microphones.

The results from the test set in which stimuli were generated from unaltered room IRs show that in halls with very different sizes and shapes that most of the differences in LEV can be attributed to level differences. This finding shows that an omnidirectional level term is sufficient as a coarse predictor for LEV. However, with smaller differences in LEV, the correlation increases by rejecting the front \( \pm 20^\circ \) and the rear \( \pm 70^\circ \), as was the case in test sets in which the overall level was equalized between stimuli. Additionally, this proposed metric benefits from including a portion of the early energy prior to 80 ms in the IR.

### 5.2 Future Work

Several items are proposed for future work. These items include: (1) gaining a better understanding of early reflections, (2) determining the JND of the proposed metric, mid-late spatial energy, \( J_s \), (3) evaluating the effectiveness of using \( J_i \) to predict LEV in more spaces, and (4) improving the spatial room IR measurement setup.

The effect of early reflections on the perception of LEV needs to be studied more thoroughly. In both the second and third LEV studies, the HLEV&LLEV IRs did not show a large change in LEV when time crossovers were varied in the early sound field, whereas for the LLEV&HLEV IRs manipulating the early sound field did change envelopment. The proposed metric does not
account for this phenomenon, and more work needs to be done to characterize the impact of early reflections on LEV.

The proposed metric should be evaluated in more listening spaces to see how it performs in a wider range of halls. Specifically, the third LEV study lacked large spaces with long reverberation times, which may impact the time integration limits. Additionally, more subjects could be included in future subjective tests to increase statistical power.

The JND of the proposed metric should be determined in order to enable meaningful comparisons between two measurements. It is also possible that there is a different JND for the overall level component of the metric than for the spatial component.

Finally, improvements could be made to the source component of the spatial room IR setup. The room IRs used in these studies were all measured with an omnidirectional source placed in the center of the stage. One improvement could be to use a sound source with a more realistic directivity pattern that is representative of musical instruments. Another improvement could be to use multiple source positions in order to generate auralizations with the instruments distributed across the stage. These improvements would add more time and complexity to the required measurements, but would enable the generation of subjective test stimuli that could sound more realistic. Whether increased realism would alter LEV perception is currently unknown.
Appendix A: **Calculation of Mid-Late Spatial Energy ($J_S$)**

This appendix documents the process of calculating mid-late spatial energy, $J_S$, from spherical microphone array data. Impulse responses (IRs) are measured using an omnidirectional sound source placed on the center of the stage with the spherical microphone array located at the listener’s seat. The spatial Fourier transform of the IRs is computed by:

$$\bar{P}_{nm}(f) = \frac{1}{b_n(f)} \frac{4\pi}{S} \sum_{s=1}^{S} P_s(f) Y_{n}^{m*}(\theta_s, \phi_s),$$  \hspace{1cm} (A-1)

where $P_s$ is the complex pressure in the frequency domain measured at microphone $s$, obtained by performing a discrete Fourier transform (DFT) on each IR, $Y_n^m(\theta, \phi)$ are spherical harmonics of order $n$ and degree $m$, * denotes complex conjugate, and $(\theta_s, \phi_s)$ is the location of the microphone on the sphere [25]. The $b_n$ coefficients for a rigid sphere are:

$$b_n = j_n(ka) - \frac{j'_n(ka)}{h_n^{(2)}(ka)} h_n^{(2)}(ka),$$  \hspace{1cm} (A-2)

where $j_n$ are spherical Bessel functions and $h_n$ are spherical Hankel functions. The inverse of $b_n$ can be applied using radial filters [72] [73]. The IRs are beamformed into a grid of directional IRs via plane wave decomposition (PWD):

$$P_{\text{grid}}(f, \theta_l, \phi_l) = \sum_{n=-N}^{N} \sum_{m=-n}^{n} \bar{P}_{nm}(f) Y_n^m(\theta_l, \phi_l),$$  \hspace{1cm} (A-3)

where $(\theta_l, \phi_l)$ is the look direction of the beam in which the main lobe of the beampattern is oriented. The directional IRs are computed every 3° in azimuth and elevation for a total of 7,200 IRs. An inverse DFT is applied to transform the IRs into the time domain. The time domain IRs are then filtered into octave bands, and $J_{SHS}$ is computed by integrating the IRs over elevation, azimuth, and time:
\[ J_s = \left( \int_{120^\circ}^{340^\circ} + \int_{120^\circ}^{340^\circ} \right) \left\{ \int_{30^\circ}^{400 \text{ ms}} \int_{30^\circ}^{60 \text{ ms}} p_{\text{grid}}(t, \theta_l, \phi_l) \sin(\theta_l) \, dt \, d\theta_l \right\} \, d\phi_l. \] (A-4)
Appendix B: Three Way Omnidirectional Sound Source

B.1 INTRODUCTION

An omnidirectional sound source is required to measure room acoustics metrics as defined by ISO 3382 [43]. Typically, the source used is a dodecahedron loudspeaker that has approximately twelve 5” drivers and has an overall diameter of about 1’ since this design has adequate sound power to excite a concert hall and radiates omnidirectionally over a wide band. However, dodecahedron loudspeakers of these approximate dimensions suffer from two disadvantages: 1) the loudspeakers typically have a low-frequency roll-off below approximately 100 Hz, and 2) the loudspeakers no longer radiate omnidirectionally above approximately 1.5 kHz. The low-frequency roll-off of a Brüel & Kjær OmniPower Type 4292 loudspeaker measured in an anechoic chamber is shown in Figure B-1, which has twelve 5-inch (13-cm) drivers and an overall diameter of 15” (39 cm). A directivity plot exhibiting lobes in the 2 kHz one-third octave band is shown in Figure B-2.

![Figure B-1: Low frequency measurement of a Brüel & Kjær OmniPower Type 4292 loudspeaker in an anechoic chamber.](image)
Since dodecahedron loudspeakers of this size exhibit lobes in the directivity response at high frequencies, ISO 3382 allows for deviation from omnidirectional directivity at high frequencies (see Figure B-3). The allowable deviation increases from ±1 dB in the 125, 250, and 500 Hz octave bands up to ±6 dB in the 4 kHz octave band. Although the measurement standard allows for this behavior, it has been shown that room acoustic measurements can vary widely based on the angular orientation of the source [89].
In order to extend the bandwidth of omnidirectional radiation, a three-way source was constructed consisting of a subwoofer, a mid-frequency dodecahedron, and a high-frequency dodecahedron (see Figure B-4). The subwoofer consists of two 10-inch (25-cm) drivers (TC Sounds Epic 10” DVC), placed on opposite sides of a 12-inch sealed cube. The mid-frequency dodecahedron contains 12 4-inch (10-cm) drivers (Tang Band W4-1658SB). The mid-frequency loudspeaker enclosure was constructed out of 16 gauge steel and has a diameter of 12 inches (30 cm). The steel was first water-jet cut from a flat sheet to make up two halves of the loudspeaker enclosure, which was then formed into the shape shown and the seams were welded together. The high-frequency dodecahedron made with 12 closely spaced 0.75-inch (1.9-cm) dome tweeters (Vifa OX20SC00-04). The high-frequency enclosure was 3D printed and has a diameter of 3 inches (8 cm). The frequency ranges for the three components are 40 Hz – 120 Hz, 120 Hz – 1.5 kHz, and 1.5 kHz to 20 kHz, respectively.
A three-way digital crossover is applied to the three IRs comprised of linear phase FIR (finite impulse response) filters which roll off at 48 dB/octave at 120 Hz and 1.5 kHz, respectively. The crossover filters were designed such that they sum to a delayed impulse in the time domain and have a flat magnitude response in the frequency domain. The frequency response of the filters is shown in Figure B-5.

*Figure B-4: Three-way omnidirectional sound source components: subwoofer (left), mid-frequency dodecahedron (middle), and high-frequency dodecahedron (right). Note that the photos are not to scale relative to each other.*

*Figure B-5: Three-way crossover filters for the omnidirectional sound source.*
B.2 DIRECTIVITY OF THE SOURCE COMPONENTS

B.2.1 Subwoofer Directivity

Free-field impulse response measurements were obtained of the subwoofer in a large open area outdoors, a concrete-paved parking lot, using the ground-plane measurement technique [82]. The frequency response of the subwoofer measured at different angles are shown in Figure B-6. The curves are normalized so that the maximum level is 0 dB. Measurements were obtained directly on-axis with the front driver and rear driver, shown in red and blue, and measurements with the source rotated 90° off-axis are shown in green and cyan. The four curves begin to diverge substantially at approximately 125 Hz, which is where the subwoofer no longer radiates omnidirectionally.

Figure B-6: Free-field measurements of the subwoofer. Measurements on-axis with the drivers are in red and blue, and measurements with the source rotated 90° off-axis are shown in green and cyan.

B.2.2 Mid-frequency dodecahedron directivity

The mid-frequency source directivity was measured in an anechoic chamber in two planes: one in which the source was rotated in azimuth, and one in which the source was rotated in elevation. In these two planes, polar directivity plots were generated by taking the averaged energy in each third-octave band from each impulse response and plotting as a function of angle. Each plot is normalized so that the maximum level is at 0 dB. Shown in Figure B-7 and Figure B-8 are directivity plots in the azimuth and elevation planes, respectively, at 500 Hz, 1 kHz,
1.6 kHz, and 2.5 kHz. The directivity of the source is omnidirectional at 1 kHz, but starting at the 1.6 kHz third-octave band the directivity begins to deviate from omnidirectional. By 2.5 kHz, distinct lobes begin to appear. The full set of directivity data for the mid-frequency dodecahedron is given in Appendix C.

Figure B-7: Directivity of the mid-frequency dodecahedron source rotated in azimuth.
B.2.3 High-frequency dodecahedron directivity

The high-frequency dodecahedron source directivity was measured in an anechoic chamber in full 3D using 18 elevation planes. Directivity plots were generated by taking the averaged energy in each third octave band from each impulse response and plotting as a function of angle. 3D balloon plots of the directivity of the source in 3D are shown in Figure B-9 at 1 kHz, 4 kHz, 5 kHz, and 6.3 kHz. The high-frequency dodecahedron begins to get directional starting at approximately 5 kHz, and distinct lobes begin to appear at 6.3 kHz. The full set of directivity data for the high-frequency dodecahedron is given in Appendix C.
B.3 OMNIDIRECTIONAL SOURCE VALIDATION WITH ROOM MEASUREMENTS

Measurements were obtained of the omnidirectional source in the Eisenhower Auditorium at Penn State University. The auditorium is a multipurpose venue used for a wide range of performance types and lectures. Four receiver locations were measured: two on the main floor (R1 and R2), one on the grand tier level (R3), and one on the balcony level (R4). The purpose of the validation measurements was to 1) evaluate the effect of source rotation when the source is no longer omnidirectional, and 2) evaluate the impact of stacking the three components as opposed to measuring the three components placed coincidently.

Figure B-9: 3D directivity plots of the high-frequency dodecahedron source in dB.
B.3.1 Measurement uncertainty due to source rotation

Measurements of the tweeter dodecahedron were obtained in which the orientation of the source was rotated in azimuth from 0° to 180° in 10° steps. The measurements were analyzed to evaluate 1) the difference in energy in the IR as a function of source rotation and 2) the difference in room acoustic metrics as a function of source rotation. All analysis and figures that follow are for R1, which reflect the worst-case differences.

B.3.1.1 Energy as a function of source rotation

The measured IRs were time windowed into early energy (0 – 80 ms) and late sound (80 ms to ∞). For each time window, a discrete Fourier transform was performed, and one-third octave smoothing was applied. The average energy across all source orientations was subtracted from each individual measurement in dB, resulting in a plot of deviations from the average energy response.

The differences in early energy are shown in Figure B-11. The differences up to 5 kHz are within ±1 dB up to 5 kHz, where the source is radiating omnidirectionally. Beyond 5 kHz, lobes appear in the directivity response of the loudspeaker, resulting in large energy differences. Above 10 kHz
these differences exceed ±10 dB. The early sound field is particularly susceptible to the
directivity of the source.

![Early energy differences as a function of source rotation in dB.](image1)

*Figure B-11: Early energy differences as a function of source rotation in dB.*

Differences in the late energy are shown in Figure B-12. In contrast to the early sound field, the
energy in the late sound field above 5 kHz is more consistent. The differences as a function of
source rotation are within approximately ±1 dB.

![Late energy differences as a function of source orientation in dB.](image2)

*Figure B-12: Late energy differences as a function of source orientation in dB.*
B.3.1.2 Room acoustic metrics as a function of source rotation

For each source rotation angle, the following metrics were evaluated as outlined in ISO 3382 [43]: early decay time (EDT), reverberation time (T30), clarity index (C80), early lateral energy fraction ($J_{LF}$) and late lateral energy level ($L_j$). For each metric, the average value across source orientation was subtracted from each measurement to compute the differences from the average. These deviations are shown in Figure B-13 through Figure B-17. In each figure, the just noticeable difference (JND) is plotted as a thick orange line. For all metrics evaluated the deviations from the average are well within 1 JND through the 4-kHz octave band. The differences exceed 1 JND in the 8-kHz octave band for EDT and C80, which are heavily dependent on the early sound field.

![Figure B-13](image)

*Figure B-13: Differences in early decay time (EDT) as a function of source orientation. JNDs for each octave band are denoted by the orange lines, where the JND for EDT is 5%.*
Figure B-14: Differences in reverberation time (T30) as a function of source orientation. JNDs for each octave band are denoted by the orange lines, where the JND for T30 is 5%.

Figure B-15: Differences in clarity index (C80) as a function of source orientation. JNDs for each octave band are denoted by the orange lines, where the JND for C80 is 1 dB.
Figure B-16: Differences in early lateral energy fraction ($J_{LF}$) as a function of source orientation. JNDs for each octave band are denoted by the orange lines, where the JND for $J_{LF}$ is 0.05.

Figure B-17: Differences in late lateral energy level ($L_J$) as a function of source orientation. JNDs for each octave band are denoted by the orange lines (the JND of $L_J$ is assumed to be 1 dB, the JND for strength, since the JND for $L_J$ is not known).

### B.3.2 Stacked vs. coincident configurations

For convenience, it would be desirable to conduct measurements in a configuration where the three components of the source are stacked one on top of the other (see Figure B-18) other as opposed to measuring the three components placed in the same location individually (see Figure B-19). The former requires one measurement, while the latter requires three separate
measurements. While convenient, the stacked configuration causes the acoustic centers of the components to be shifted vertically, thus not all in the same location. Additionally, the sound scattered off adjacent components modifies the source radiation directivity and frequency response. IRs were measured in both configurations to determine whether measurements obtained in the stacked configuration are valid.

Figure B-18: Three-way omnidirectional source in the stacked configuration.

Figure B-19: Three-way omnidirectional source in the coincident configuration, three separate measurements.
B.3.2.1 Energy differences between stacked and coincident configurations

The measured IRs were time windowed into early energy (0 – 80 ms) and late sound (80 ms to ∞). For each time window, a discrete Fourier transform was performed, and one-third octave smoothing was applied. The energy in the coincident configuration was subtracted from the energy in the stacked configuration to evaluate the differences in dB.

The differences in the early energy between the two configurations are shown in Figure B-20. The worst case differences between the two configurations are shown, which again were found in R1. The stacked configuration causes errors up to 6 dB in the subwoofer, 2 dB in the mid-frequency dodecahedron, and 10 dB in the high-frequency dodecahedron. These differences are unacceptably large for all three components.

![Figure B-20: Difference in early energy between the stacked configuration and coincident configuration.](image)

The differences in late energy between the two configurations are shown in Figure B-21. The differences in the late sound field are in general smaller than the differences in the early sound field. However, the differences between the two configurations is still unacceptably large. The differences are up to 4 dB in all three components.
B.3.2.2 Room acoustic metrics for the stacked and coincident orientations

For the two configurations, the following metrics were evaluated as outlined in ISO 3382 [43]: early decay time (EDT), reverberation time (T30), clarity index (C80), and late lateral energy level (L_l). For each metric, the coincident configuration value was subtracted from the stacked configuration value to yield differences in configuration. These differences are shown in. In each figure, the just noticeable difference (JND) is plotted as a thick orange line. For all metrics evaluated except for T30, the differences exceed 1 JND. These results suggest that the stack configuration should not be used to evaluate room acoustics metrics, since the interaction between sources causes a large impact in the measurements.
Figure B-22: Differences in early decay time (EDT) for the stacked vs. coincident configurations. JNDs for each octave band are denoted by the orange lines, where the JND for EDT is 5%.

Figure B-23: Differences in reverberation time (T30) for the stacked vs. coincident configurations. JNDs for each octave band are denoted by the orange lines, where the JND for T30 is 5%.
Figure B-24: Differences in clarity index \(C^{80}\) for the stacked vs. coincident configurations. JNDs for each octave band are denoted by the orange lines, where the JND for \(C^{80}\) is 1 dB.

Figure B-25: Differences in late lateral energy level \(L_J\) for the stacked vs. coincident configurations. JNDs for each octave band are denoted by the orange lines (the JND of \(L_J\) is assumed to be 1 dB, the JND for strength, since the JND for \(L_J\) is not known).

**B.4 SUMMARY**

A three-way omnidirectional sound source was constructed consisting of a subwoofer, a mid-frequency dodecahedron, and a high-frequency dodecahedron. The subwoofer extends the low frequency bandwidth of the omnidirectional source down to 40 Hz, whereas conventional
omnidirectional sources experience a roll-off below approximately 100 Hz. The high-frequency dodecahedron extends the omnidirectional directivity at high frequencies up to 5 kHz, whereas conventional dodecahedrons are omnidirectional up to approximately 1.5 kHz. Room acoustics metrics were evaluated for the source as a function of orientation, and it was determined that the measurements are largely unaffected by the source orientation through the 4-kHz octave band. A measurement configuration was evaluated in which the three components of the system were stacked for ease the measurements, and large differences were found when compared to a measurement configuration in which the three components were placed coincidently and measured separately. For this reason, the stacked configuration is not recommended.
Appendix C: Complete Set of Directivity Plots of the Mid-Frequency and High-Frequency Dodecahedrons

C.1 HIGH-FREQUENCY DODECAHEDRON: 3D OCTAVE BAND PLOTS (1,000 – 16,000 Hz)
C.2 HIGH-FREQUENCY DODECAHEDRON 3D THIRD-OCTAVE BAND PLOTS (1,000 – 16,000 HZ)
C.3 HIGH-FREQUENCY DODECAHEDRON: 2D POLAR PLOTS (AZIMUTHAL PLANE) THIRD-OCTAVE BANDS (1,000 – 20,000 HZ)

- 1000 Hz
- 1260 Hz
- 1587 Hz
- 2000 Hz
- 2520 Hz
- 3175 Hz
C.4 HIGH-FREQUENCY DODECAHEDRON: 2D POLAR PLOTS (0° ELEVATION PLANE) THIRD-OCTAVE BANDS (1,000 – 20,000 Hz)
C.5 HIGH-FREQUENCY DODECAHEDRON: 2D POLAR PLOTS (90° ELEVATION PLANE) THIRD-OCTAVE BANDS (1,000 – 20,000 HZ)
C.6 MID-FREQUENCY DODECAHEDRON: 2D POLAR PLOTS (AZIMUTHAL PLANE) THIRD-OCTAVE BANDS (125 – 20,000 HZ)
C.7 MID-FREQUENCY DODECAHEDRON: 2D POLAR PLOTS (ELEVATION PLANE) THIRD-OCTAVE BANDS (125 – 20,000 HZ)
Appendix D: **Subjective Test Tutorial Example**

This appendix documents an example tutorial given to subjects at the beginning of each listening test. This particular example is for the subjective test in Ch. 4. The tutorials for all the subjective listening tests described in this dissertation follow the same format, with differences only for the specific formats in each test (e.g. test set durations and specific GUI instructions).
Purpose of Research

- The purpose of this research is to determine the factors which influence listener envelopment in concert halls.

- The results of this study may be helpful to the architectural acoustics community in the future research and design of performing arts spaces.

Subjective Testing

- This testing is a very important phase of the research, as we use it to see whether listeners can detect differences in musical tracks when variables in the hall are changed.

- Your participation is greatly appreciated!
Test Structure

- During the testing you will listen to:
  - 1 practice set containing 4 test questions
  - 1 set containing 6 test questions
  - 4 sets containing 21 or 25 test questions each

- You will take a 5 minute break after approximately 15 minutes of testing

- The sets will be followed by a final wrap-up and paperwork

Instructions

The tracks in this test will contain short clips of music from an orchestra:
Spatial Impression and LEV

While listening, pay attention to what you hear spatially. The spatial impression of a concert hall can be divided into two components:

1. **Apparent source width (ASW)**
   The sense of how *narrow* or *wide* the source appears on the stage.

2. **Listener envelopment (LEV)**
   The sense of feeling *surrounded* by the sound or *immersed in* the sound.

---

Spatial Impression and LEV

- During this test, we ask that you **only** focus on the perception of **Listener Envelopment**, the sense of feeling *surrounded* by the sound or *immersed in* the sound.
- Please try to ignore all other aspects of the sound field, including **Apparent Source Width**, and reverberance.
Instructions

During the test, you will be asked to listen to two orchestral audio clips, A and B. Choose the sound field which is more \textit{enveloping} (i.e. surrounding, immersive).

You are encouraged to switch back and forth as many times as necessary to make your judgement.

Instructions

Next, you will be asked to rate how much more enveloping your selection is, compared to the other clip, on a scale from 0 to 10 using the slider.

- A rating of 10 means that your selection is \textbf{EXTREMELY MORE ENVELOPING}.
- A rating of 0 means that A and B are \textbf{EQUALLY ENVELOPING}.
Instructions

- The testing will be done with a computer program. You will listen to and rate each presented track.

- Your responses will be recorded using a tablet interface. Answers can be recorded either using the touch screen on the tablet, or a mouse, whichever you prefer.

- The following slides will introduce you to the program.

User Interface

- Below is a picture of the full user interface.
- Switch between A and B by clicking the blue “Play A” and “Play B” buttons, respectively.
User Interface

• Click the Mute button to turn the audio off.

User Interface

• The progress bar shows how far along you are in the excerpt.
• To loop over a small section of the excerpt, click and drag over the desired section on the purple bar, and click “Limit Passage.”
User Interface

> Click “Full Passage” to return to the full excerpt.

"Restart Passage" will start the excerpt from the beginning.
User Interface

- Make your selection by clicking the red buttons to indicate whether sound field A or B is more enveloping.
- If they are equally enveloping, select either one.

User Interface

- After selecting A or B, the rating slider will appear.
- Rate each excerpt by dragging the top of the bar up or down to the desired rating.

![Diagram of User Interface](image-url)
User Interface

- Once you are satisfied with the ratings of the excerpt, click “Submit Rating” to move on to the next question.
- You are not required to listen to the entire passage.

Instructions

While listening, imagine yourself in the audience of a concert hall, listening to an orchestra that is in front of you.
Final Notes

- You are encouraged to rotate your head during the test. This may help you to hear the spatial differences between the excerpts.
- You must keep your head between the two copper discs marking the center of the loudspeaker array, and keep your ears at the height of the discs.
- If your head moves too far from the center position, you will be asked to reposition.

Final Notes

- Dave will be able to see and hear you throughout the entire test. If you need to grab his attention for any reason, just say so.
- Try not to hold your finger down in one place on the tablet, since the touch screen will register it as a right click.
- Thank you very much for participating in the study!
References


Vita

David A. Dick

David Dick grew up in Norwalk, CT and graduated from Norwalk High School in 2006. As a guitar player, David’s interest in music led him to study acoustics at the University of Hartford. In 2010, he graduated *Summa Cum Laude* with a Bachelor of Science degree in Acoustical Engineering and Music (jazz guitar concentration), a Bachelor of Science degree in Electrical Engineering, and a minor in Mathematics. After graduating, David worked at Bose Corporation in Framingham, MA as an acoustical engineer while taking distance education coursework in the Graduate Program in Acoustics at The Pennsylvania State University. In 2013, he transitioned to a full time resident student to pursue a Doctor of Philosophy degree researching concert hall acoustics with Dr. Michelle Vigeant. David currently works at Bose Corporation in the advanced development group in the Consumer Electronics Division. His professional interests include audio signal processing, spatial audio, psychoacoustics, and loudspeaker design.