INVESTIGATION OF LISTENER ENVELOPMENT AND THE LATE SOUND FIELD USING SPHERICAL MICROPHONE ARRAY IMPULSE RESPONSE MEASUREMENTS

D A Dick The Graduate Program in Acoustics, The Pennsylvania State University, University Park, Pennsylvania, USA
M C Vigeant The Graduate Program in Acoustics, The Pennsylvania State University, University Park, Pennsylvania, USA

1 INTRODUCTION

The purpose of this study was to investigate how the perception of listener envelopment (LEV), the sense of being immersed in a sound field, relates to the late sound field utilizing impulse response (IR) measurements obtained with a spherical microphone array. Spherical microphones are composed of a number of microphone elements arranged on the surface of a sphere, and can be used to beamform directional patterns to obtain spatial information about sound fields. Current LEV metrics, such as Late Lateral Energy Level (L_J), are based on measurements of lateral reflections, which are typically acquired with a figure-of-eight microphone.[1] Spherical microphone arrays allow for a higher resolution analysis than traditional methods, and can be used to analyze the sound field over full 3D space. Additionally, the spherical array measurements can be reproduced over a loudspeaker array using Ambisonics in order to run subjective studies.

In this study, IR measurements were obtained in the Peter Kiewit Concert Hall in Omaha, NE, USA with an Eigenmike em32 spherical microphone array. The spherical array IRs were used for both an objective sound field analysis, and for reproduction over a loudspeaker array for subjective testing. Listening tests were conducted using third-order Ambisonic reproductions of the measured IRs convolved with anechoic music recordings played over a 30-loudspeaker array in an anechoic chamber. The late sound field was analyzed objectively and compared to the subjective test results.

2 BACKGROUND INFORMATION

2.1 Listener Envelopment

Early research quantifying the sense of spatial impression in concert halls began in the 1960's and has continued to present day. Initial subjective characteristics were referred to generally as 'spatial responsiveness', 'spatial reverberation', and 'spatial impression'.[2] The earliest metrics that were used to quantify the spatial impression of a room were: Early Lateral Energy Fraction (J_{LF}), the ratio between the early lateral sound in the first 80 ms and the total early sound[3] and Interaural Cross Correlation Coefficient (IACC), found from the cross-correlation of the left and right ears of a binaural measurement. The overall spatial impression of a room has since been shown to contain two distinct perceptions: Apparent Source Width (ASW), the sense of how wide or narrow the sound image appears to the listener, and Listener Envelopment (LEV), the sense of being immersed in and surrounded by the sound field.[4,5] ASW has been found to be a function of early lateral reflections, which causes a perceptual widening of the sound source.[3] The perception of LEV has been found to be a function of late lateral sound. [6] Since LF and IACC are related to the early reflections, these metrics have been shown to correlate with the perception of ASW.

To begin to investigate of the perception of LEV, a study was conducted that evaluated the different components of the sound field.[6] The IRs were simulated in an anechoic chamber using five loudspeakers distributed azimuthally in the horizontal plane. A limited number of early room reflections were played out of five individual loudspeakers placed in the frontal half of the median plane and were kept constant throughout the study. The late sound field was modified by changing the reverberation time (T30), the ratio of early-to-late energy – quantified using clarity index (C80), the overall late sound level (G_{Late}), and the angular distribution of the late sound. To vary the angular distribution, the late sound was played out of either a single loudspeaker directly in front of the listener, three loudspeakers spanning 70° in front of the listener, or five loudspeakers spanning 180° in front of the listener. While the authors found a weak correlation in all of the IR properties that were varied, the findings indicated that the largest contributors to LEV were angular distribution and overall level. The LEV ratings obtained from the listening test were evaluated using several proposed objective metrics, and the authors introduced a metric called Late Lateral Energy Level, (L_J, prior notations GLL, LG, or LG_{80}^{∞}):[1]

$$L_J = 10\log_{10}\left[\frac{\int_{80ms}^{\infty} p_L^2(t)dt}{\int_0^{\infty} p_{10}^2(t)dt}\right] \text{ [dB]}, \qquad (1)$$

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. This metric had the highest correlation with LEV ratings out of all of the metrics evaluated in the study. To date, L_J is the most commonly used metric to predict the sense of listener envelopment and is included in the annex of the most recent version of the ISO room acoustics parameters standard. [7]

Aside from L_J, other metrics have been proposed to objectively measure LEV. The Late Interaural Cross Correlation Coefficient ($IACC_{L,3}$) is a metric calculated from the cross-correlation between the left and right ears of a binaural IR measurement from 80 milliseconds to 1 second in the 500 Hz to 2 kHz octave bands. This parameter has very little variation between halls and has been found to be a poor predictor of LEV.[8] Late Lateral Energy Fraction (LLF), the ratio of late lateral sound to total late sound, has also not been found to significantly vary between halls or within halls, and thus is not a good predictor of LEV.[9] A formula to calculate LEV objectively based on $IACC_{L,3}$, Strength (G) and Clarity Index (C80) was proposed by Beranek.[10] In 2001, a metric called Spatially Balanced Center Time (SBT_S) was introduced based on the center time of the IRs weighted by arrival direction.[11] It should be noted that none of these metrics have gained any traction in the architectural acoustics community, and there are only a limited number of studies that evaluate the performance of each of these metrics.

2.2 Spherical Microphone Array Processing and Beamforming

The sound pressure on the surface of a sphere due to an incident plane wave can be represented as an infinite sum of spherical harmonics:[12]

$$p(r,\vartheta,\varphi,t) = P_0 4\pi \sum_{n=0}^{\infty} i^n b_n(ka) \sum_{m=-n}^{n} Y_n^m(\vartheta,\varphi) Y_n^{m*}(\vartheta_i,\varphi_i)\, e^{i\omega t}, \qquad (2)$$

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

$$Y_n^m(\vartheta,\varphi) = \sqrt{\frac{(2n+1)}{4\pi}\frac{(n-m)!}{(n+m)!}} P_n^m(cos\vartheta) e^{im\varphi}. \qquad (3)$$

The coefficients b_n are referred to as plane wave modal coefficients, and for a rigid sphere, the coefficients are:[13]

$$b_n = j_n(kr) - \frac{j_n{}'(ka)}{h_n^{(2)'}(ka)} h_n^{(2)}(kr), \qquad (4)$$

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, and therefore, 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:

$$\tilde{P}_{nm}(ka) = \frac{1}{b_n(ka)} \sum_{s=1}^{S} P_s\,(ka) Y_n^{m*}(\vartheta_s, \varphi_s), \quad (5)$$

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 (ϑ_s, φ_s) is the location of the microphone on the sphere. The spatial Fourier components can be weighted and combined to perform a plane wave decomposition (PWD) [14]:

$$P(\vartheta_l, \varphi_l) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \tilde{P}_{nm}(ka) Y_n^m(\vartheta_l, \varphi_l)\,, \quad (6)$$

where (θ_l, ϕ_l) is the look direction of the beam, or the direction in which the maximum of the beampattern is oriented. Directional IRs can be obtained by performing a PWD on the IRs measured with a spherical array.

2.3 Ambisonics Reproduction

Ambisonics is a surround-sound reproduction system developed by Gerzon in 1973 as a method to reproduce sound fields represented in the spherical harmonics domain.[15] Ambisonics can be thought of as sampling the spherical harmonic basis functions in space with loudspeakers. In traditional Ambisonic reproduction, the zeroth-order (monopole) and first-order (dipole) spherical harmonic components are reproduced over a loudspeaker array, yielding a recreation of the acoustic pressure and particle velocity at a single listening point. Ambisonics has since been extended to higher orders, referred to as Higher Order Ambisonics (HOA). In HOA, additional spherical harmonic components are included, which has the effect of widening the effective listening area or “sweet spot”.

3 ROOM IMPULSE RESPONSE MEASUREMENTS

IR measurements were made in the Peter Kiewit Concert Hall in Omaha, NE, USA. The Peter Kiewit Concert Hall features variable absorption in the form of absorptive panels on the ceiling and walls which can be either retracted or deployed. Three hall settings were measured with mid-frequency average reverberation times of 1.8, 2.4, and 2.8 seconds. For each of these settings, 10 receiver locations were measured throughout the hall, shown in Figure 1. The measured IRs were used both for Ambisonics reproduction in a subjective study and for objective analysis of the sound field using PWD.

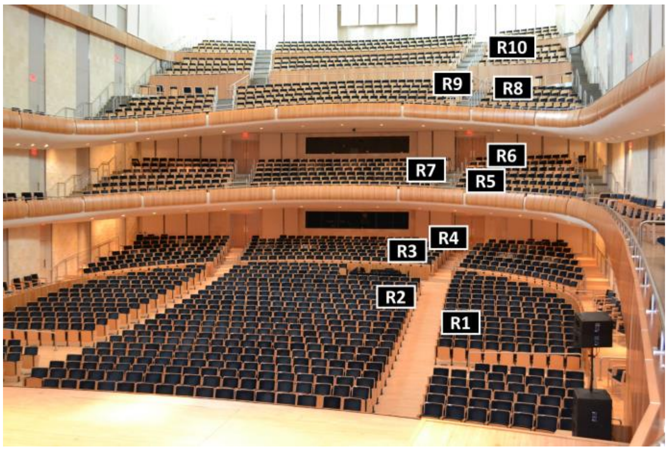


Figure 1: Receiver locations in the Peter Kiewit Concert Hall.

The IR measurements were obtained with the room acoustics software EASERA[16] using the 2-channel FFT correlation measurement technique. A logarithmic sine sweep excitation was used with eight averages for each measurement. The sound source for the IR measurements was a Brüel & Kjær Type 4292 dodecahedron loudspeaker placed on the center of the stage, shown in Figure 2 (a). Measurements were made using an Eigenmike spherical microphone array made by mh acoustics, shown in Figure 2 (b).[17] The Eigenmike is a rigid sphere with a diameter of 8.4 centimeters containing 32 omnidirectional microphones spaced according to the center of the faces on a truncated icosahedron. This sampling scheme preserves the orthogonality property of the spherical harmonics up to third order. Additional measurements were also made using a Brüel & Kjær Type 4100-D binaural mannequin as shown in Figure 2 (c).

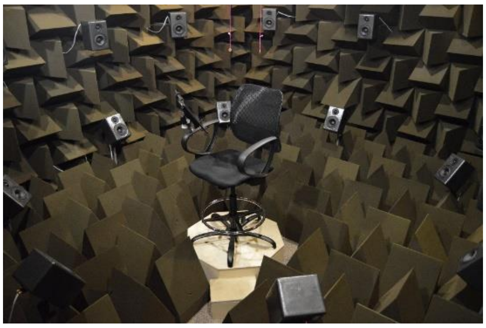
(a) (b) (c)

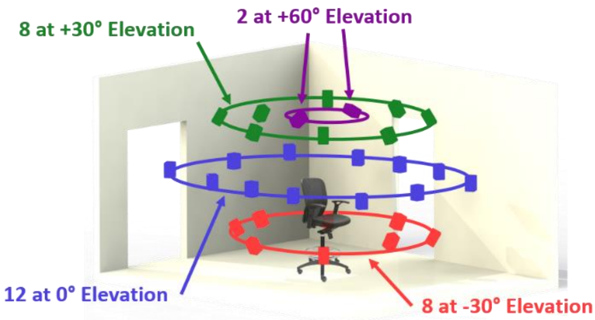
Figure 2: Measurement hardware used for IR measurements: B&K Dodecahedron loudspeaker (a), Eigenmike spherical microphone array (b), and B&K binaural mannequin (c).

4 LISTENER ENVELOPMENT SUBJECTIVE TEST

4.1 Ambisonics Reproduction for the Subjective Study

For the subjective study, Ambisonics reproduction was accomplished over a 30-channel loudspeaker array in a nearly-spherical distribution in the AUralization Reproduction of Acoustic Sound fields (AURAS) facility at Penn State, as shown in Figure 3. The loudspeakers are placed in three rings: 8 loudspeakers at 30 degrees below the horizontal plane, 12 loudspeakers at the horizontal plane, and 8 loudspeakers at 30 degrees above the horizontal plane. Two additional loudspeakers are located almost overhead at 60 degrees above the horizontal plane. The design and construction of the loudspeaker array is detailed in an M.S. thesis.[18] The Ambisonic decoder for this array was designed using Heller's Ambisonic Decoder Toolbox.[19,20] The decoder uses phase-matched shelf filters to cross over from basic decoding to max-r_E decoding at 400 Hz, which reduces side lobes and improves the spatial perception at high frequencies.[21] In addition, the decoder includes time delay compensation and a magnitude correction to account for the distance from each loudspeaker to the center of the array, and order-dependent high-pass filters for near field compensation.[22]





(a) (b)

Figure 3: A picture of the AURAS loudspeaker array (a) and the distribution of the loudspeakers (b).

To reproduce the measured IRs using HOA, the Eigenmike IRs were transformed into the spherical harmonics domain. Radial filters were then applied, which invert the plane wave modal coefficients in Eqn. 4.[23] These filters also included a random-incidence correction to equalize the frequency response of the Eigenmike, which has a high frequency roll off characteristic. The signals were then sent through the Ambisonic decoder to generate loudspeaker signals. The loudspeaker signals were filtered to equalize the frequency response of the array's individual loudspeakers, as well as the frequency response of the omnidirectional loudspeaker used to measure the IRs. The signal processing in this study was performed using VST plugins from Kronlacker's ambiX and mcfx plug-in suites[24] in the digital audio workstation software REAPER.[25]

The Ambisonics reproduction of the stimuli used in this study has been evaluated using informal ABX listening tests. Binaural recordings of the reproduction in the array were recorded and compared to binaural recordings of the original sound field. The subjective differences were nearly inaudible. Additionally, broadband plane waves were generated and panned around in 3D space to ensure proper sound localization in all directions.

4.2 Subjective Study Details

A subjective study was run using the IR measurements described in Section 3. Participants were asked to rate stimuli in terms of perceived envelopment. The room IRs measured with the Eigenmike spherical array were processed for spatial playback as described in Section 4.1 and were convolved with an anechoic music excerpt, Bizet's L'Arlesienne Suite No. 2.[26] The stimuli were presented in four sets of eight signals in a completely randomized experimental design. Set 1 contained IRs from the most absorptive setting, Set 2 contained IRs from the most reverberant setting, Set 3 contained IRs from a setting in-between the most absorptive and most reverberant settings. Set 4 contained a mixture of IRs from the three aforementioned settings that all had similar L_J values.

In 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 graphical user interface (GUI) shown in Figure 4. 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.

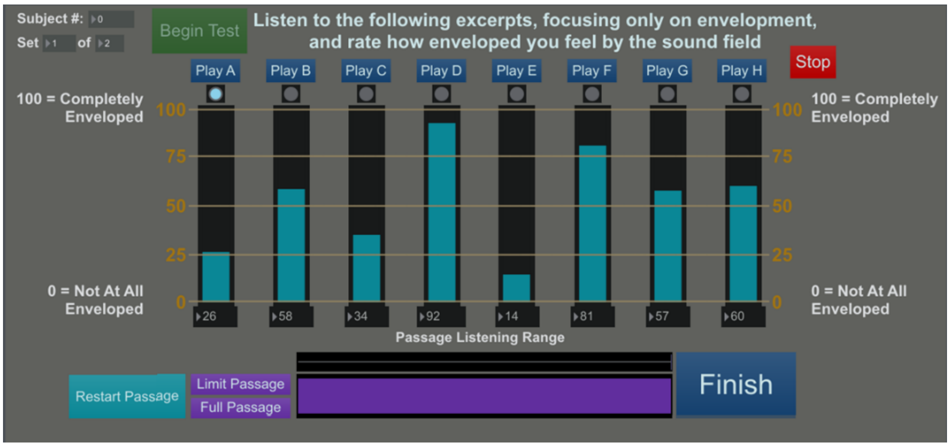


Figure 4: LEV subjective testing GUI enabling instantaneous switching between eight stimuli.

The subjective study was run with 15 test subjects, all with minimum hearing thresholds of 15 dBHL from 250 Hz to 8000 Hz, which were measured. Additionally, all subjects had a minimum of 5 years of formal musical training and were musically active (i.e. performing in an ensemble and/or taking private music instruction) at the time of the study.

5 RESULTS AND DISCUSSION

5.1 Subjective Test Results

A one-way repeated-measures analysis of variance (ANOVA) was run on the subjective test data. Sets 1 through 3 yielded significant p-values ($p < 0.001$, $p < 0.001$, and $p = 0.014$, respectively), indicating that these sets contain significant differences in the mean LEV ratings between some of the receiver positions. Set 4 was insignificant ($p = 0.071$), likely due to the low number of subjects. For the three sets with significant p-values, pairwise t-tests were conducted to find pairs with significant differences. Within Sets 1 and 2, pairs were identified in which the LEV ratings were significantly different but values of L_J were similar. Conversely, within the same sets pairs were found in which the L_J values were significantly different but LEV ratings were found to be similar. For each of these cases, an example pair is listed in Table 1. In order to examine the relationship between LEV ratings from Sets 1 through 3 and the 3D late sound field, the IRs were analyzed using PWD (Eqn. 6) as described in section 5.2.

Table 1: Example of two pairwise comparisons: one with similar LEV ratings and slightly different L_J values (top), and one with similar L_J values and significantly different LEV ratings (bottom).

Receiver	L_J Average [dB] (125 Hz to 1 kHz)	LEV Rating	ΔL_J Average [dB]	ΔLEV Rating
Set 2 Receiver 3	-1.4	73.3	1.3[a]	1.0
Set 2 Receiver 8	-0.1	72.3		
Set 1 Receiver 9	-3.3	65.6	0.2	17.7
Set 1 Receiver 10	-3.6	48.0		

[a] The just noticeable difference (JND) of L_J is not known. It may be reasonable to assume that the L_J JND is similar to the JND of strength (G), which is 1 dB.[7]

5.2 Analysis of the Sound Fields

The spherical-array IRs measured were analyzed using PWD (Eqn. 6). The SOFiA toolbox in MATLAB was utilized for some of the spherical microphone array processing.[27] Directional IRs were generated for look-directions spaced every 3 degrees in azimuth and elevation, resulting in a grid of 60 (in elevation) by 120 (in azimuth) IRs. The IRs were then windowed in the time domain from 80 ms until the end of the IR to isolate the late sound, and subsequently octave-band filtered. The energy at each grid point was calculated by squaring and summing the windowed and filtered IRs. To visualize the late sound field, the energy was plotted both on a sphere, and on a flattened rectangle of elevation vs. azimuth (similar to a 2D map being a flattened version of a globe) using a decibel scale. In these plots, 0° azimuth refers to the front of the array pointing at the stage, and 0° elevation points straight up. Two comparisons of the late sound field analyzed using PWD are shown in Figs. 5 and 6, respectively as examples. In each comparison plot, the sound fields are normalized such that the maximum level of both plots is set to 0 dB in order to preserve the absolute differences in level between the two sound fields. The 1 kHz octave bands are shown, although trends are consistent from the 500 Hz through 4 kHz octave bands.

In Set 2, Receivers 3 and 8 both had similar LEV ratings (73.3 and 72.3, respectively), yet the late sound fields look very different as shown in Figure 5. Additionally, Receiver 8 had a value of L_J which was approximately 1.3 dB higher than Receiver 3. At Receiver 3, the late sound field is concentrated toward the front, and the energy on the sides and back are reduced, whereas at Receiver 8 the energy is more evenly distributed throughout the sphere. This finding is counter to the expectation that the late sound fields be similar since the LEV ratings are similar.

In Set 1, Receiver 9 was found to be significantly more enveloping than Receiver 10 (65.6 and 48.0, respectively, $p = 0.020$), although the differences in the late sound field between these receiver positions are relatively small, as shown in Figure 6. The average L_J values between these two receiver positions are also identical as shown in Table 1. Given that the LEV ratings are significantly different, it was expected that the late sound fields would have spatial differences, but the differences in this comparison are much smaller than the differences shown in Figure 5.

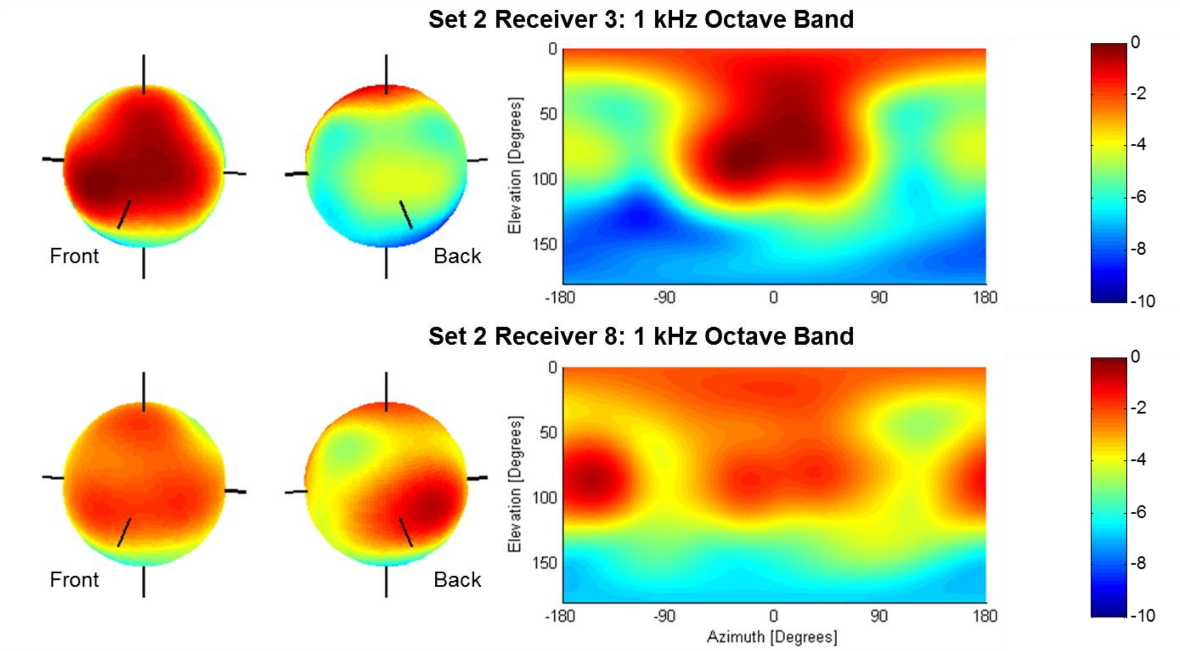


Figure 5: Comparison of the late sound field between receiver positions with similar LEV ratings, with 0° azimuth pointing toward the stage and 0° elevation pointing straight up. Energy at Receiver 3 is concentrated toward the front, whereas energy at Receiver 8 more evenly distributed throughout the sphere.

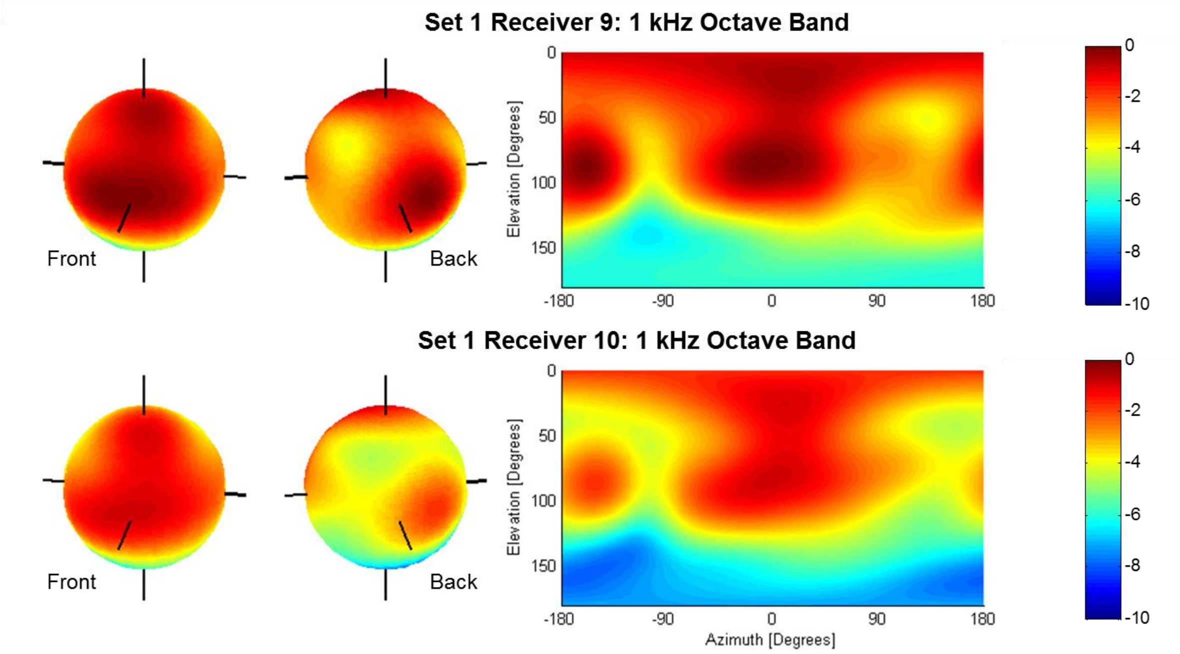


Figure 6: Comparison of the late sound field between receiver positions with different LEV ratings, with 0° azimuth pointing toward the stage and 0° elevation pointing straight up. The spatial distribution of late energy is similar between the two receivers despite significantly different LEV ratings.

6 CONCLUSIONS

In this study, the perception of listener envelopment (LEV) was investigated using spherical microphone array IR measurements obtained in the Peter Kiewit Concert Hall in Omaha, NE, USA. The measurements were used to reproduce the measured sound field over a loudspeaker array for a subjective listening test in which 15 participants rated the LEV of each sound field. Significant differences in mean LEV ratings were found in three of the four sets of stimuli used in the subjective test. In order to examine the relationship between the subjective test results and the late sound field, plane wave decomposition (PWD) was used to analyze the spatial distribution of the late energy of the IRs. This analysis revealed instances where there were substantial differences in the late sound field, but the LEV ratings were similar. Conversely, in some cases there were significant differences in LEV ratings, but similar late sound fields.

The results of the investigation of the late sound field do not indicate a significant trend between the spatial distribution of the late sound and LEV ratings, however, more studies are need to substantiate this finding. Future work will examine different time windows of the spatial IRs (i.e. the early sound field or smaller time segments) to investigate how the different components of the IRs relate to LEV. These findings will be used to develop a new metric to predict LEV based on the spherical microphone array measurements.

7 ACKNOWLEDGEMENTS

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 Lily Wang, Matt Blevins, Laura Brill, Hyun Hong, Joonhee Lee, and Ellen Peng of University of Nebraska-Lincoln for their assistance in obtaining the IR measurements. This work was sponsored by the National Science Foundation (NSF) award #1302741.

8 REFERENCES

1. Bradley, J. S. & Soulodre, G. A., Objective measures of listener envelopment. *J. Acoust. Soc. Am.* **81** (4), 344-355 (1995).
2. Marshall, A. H. & Barron, M., Spatial responsiveness in concert halls and the origins of spatial impression. *Applied Acoustics* **62** (2), 91-108 (2001).
3. Barron, M. & Marshall, A. H., Spatial Impression due to Early Lateral Reflections in Concert Halls: The Derivation of a Physical Measure. *J. Sound and Vib.* **77** (2), 211-232 (1981).
4. Morimoto, M., Fujimori, H. & Z., M., Discrimination between auditory source width and envelopment. *J. Acoust. Soc. Jpn.* **46** (6), 449-457 (1990).
5. Morimoto, M. & Maekawa, Z., *Auditory spaciousness and envelopment*, Proc. Int. Cong. on Acoust., Belgrade, 1989.
6. Bradley, J. S. & Soulodre, G. A., The influence of late arriving energy on spatial impression. *J. Acoust. Soc. Am.* **98** (5), 2263-2271 (1995).
7. *ISO 3382-1:2009: Acoustics – Measurement of room acoustic parameters – Part 1: Performance spaces*.
8. Beranek, L., *Concert Halls and Opera Houses: Music, Acoustics, and Architecture* (Springer, New York, 2003).
9. Barron, M., Late lateral energy fractions and the envelopment question in concert halls. *Appl. Acoust.* **62** (2), 185-202 (2001).
10. Beranek, L. L., *Listener Envelopment LEV, Strength G, and Reverberation Time RT in Concert Halls*, Proc. Int. Cong. on Acoust., Sydney, Australia, 2010.
11. Hanyu, T., A new objective measure for evaluation of listener envelopment focusing on the spatial balance of reflections. *Applied Acoustics* **62**, 155-184 (2001).

12. Williams, E. G., *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic Press, New York, 1999).
13. Meyer, J. & Elko, G., *A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield*, Proc. IEEE Int. Conf. on Acoust., Speech, and Signal Proc., 2002.
14. Rafaely, B., Plane-wave decomposition of the sound field on a sphere by spherical convolution. *J. Acoust. Soc. Am.* **116** (4), 2149-2457 (2004).
15. Gerzon, M., Periphony: With-Height Sound Reproduction. *J. Audio Eng. Soc.* **21** (1), 2-10 (1973).
16. EASERA v 1.2, http://easera.afmg.eu/ (Accessed 18 September 2015).
17. acoustics, m., em32 Eigenmike microphone array release notes (v17.0), http://www.mhacoustics.com/sites/default/files/ReleaseNotes.pdf (Accessed 18 September 2015).
18. Neal, M., in *Investigating the sense of listener envelopment in concert halls using a third-order ambisonic reproduction over a loudspeaker array and a hybrid room acoustics simulation method* (M.S. Thesis, Penn State University, 2015).
19. Heller, A. J., Benjamin, E. M. & Lee, R., *A Toolkit for the Design of Ambisonic Decoders*, Proc. Linux Audio Conference, Stanford, CA, 2012.
20. Heller, A. J. & Benjamin, E. M., *The Ambisonic Decoder Toolbox: Extensions for Partial-Coverage Loudspeaker Arrays*, Proc. Linux Audio Conference, Karlsruhe, Germany, 2014.
21. Zotter, F. & Frank, M., All-round ambisonic panning and decoding. *L. Aud. Eng. Soc.* **60** (10), 807-720 (2012).
22. Daniel, J., *Spatial Sound Encoding Including Near Field Effect: Introducing Distance Coding Filters and a Viable, New Ambisonic Format*, Proc. AES 23rd International Conference, Copenhagen, Denmark, 2003.
23. Baumgartner, R., Pomberger, H. & Frank, M., *Practical Implementation of Radial Filters for Ambisonic Recordings*, Proc. ICSA, Detmold, Germany, 2011.
24. Kronlachner, M., *Plug-in Suite for Mastering the Production and Playback in Surround Sound and Ambisonics*, Proc. AES Student Design Competition, Berlin, Germany, 2014.
25. Reaper Digital Audio Workstation, www.reaper.fm (Accessed 18 September 2015).
26. Denon, *Anechoic Orchestral Music Recordings,* Denon Records (1995).
27. Bernschütz, B., Pörschmann, C., Spors, S. & Weinzierl, S., *SOFiA Sound Field Analysis Toolbox*, Proc. ICSA Int. Conf. on Spatial Audio, Detmold, Germany, 2011.