VALIDATION OF THE BINAURAL ROOM SCANNING METHOD FOR CINEMA AUDIO RESEARCH

LA Gedemer University of Salford, Salford, UK
T Welti Harman International, Northridge, CA, USA

1 INTRODUCTION

Conducting subjective listening tests in situ within a commercial cinema or professional dub-stage can prove logistically difficult. Binaural Room Scanning (BRS) affords many practical, logistical and methodological benefits over in situ listening tests, including the elimination of sighted biases related to the objects being tested [1]. BRS provides the benefit of allowing larger samples of listeners to participate in testing, something that professional dub stages and public cinemas cannot easily accommodate. BRS also allows different experimental variables to be easily manipulated and presented to listeners over different time periods to study how they learn and adapt to different cinema acoustics and response curves [2].

Current cinema audio research has focused on room calibration standards, such as SMPTE ST202:2010. However, to date, the performance of the SMPTE X-curve (or an alternative calibration curve) has never been validated based on published results from controlled listening tests. The following questions serve as a basis for future cinema-related subjective listening tests:

- Do listeners’ perceptions of spectral balance change with venue size?
- What is the preferred spectral balance (in-room target response curve)
  - For conventional music tracks
  - For X-curve monitored film sound tracks?
  - For non-calibrated (flat) film sound tracks?

In order to validate the use of BRS for such future tests, a series of listening tests using trained listeners was conducted in situ using two rooms of different sizes, both with 5.1 audio playback systems. The same listening tests were repeated using binaural room scans of these rooms, and the listener responses from both of these tests were analyzed to determine their similarity. The BRS system has shown a significant positive correlation with responses to in situ listening tests for loudspeaker preference [2] and automotive audio tests [3]. Concerns regarding the contextual effects of the listening environment when comparing BRS and in situ playback have shown to be unwarranted in similar tests for automotive audio [4]. It is therefore proposed that BRS can serve as a valid tool for cinema audio research, which this study attempts to confirm. In Section 2, the main research question is described in addition to the implementation of the BRS system validation experiment. Section 3 discusses statistical analysis of the results, with discussion and conclusions in sections 4 and 5.

2 BRS SYSTEM VALIDATION

The performance target for the BRS system is relatively straightforward: the system should accurately capture a listening space and deliver the identical signals to the listeners' ears as if they had been receiving those signals in situ.
Section 2 describes the main research question and details the implementation of the BRS system validation experiment.

2.1 Research Question

The method of validating the BRS system was chosen due to its relationship to how the authors ultimately intend to use it. Its primary application will be for conducting research into cinema audio for varying sized playback spaces and systems. This experiment addresses how well the BRS system performs when used as a substitute for in situ listening, specifically when room response curves are in question. The main research question is: "Are there significant differences in listeners' room response curve preferences using BRS playback compared to in situ playback?"

2.2 Binaural Room Scanning

Two rooms were used in the investigation. One room, the “Living Room” with dimensions of 2.4 m.(H) x 4.3 m. (W) x 6.5 m. (L), approximates the size and acoustical properties of the average home living room. There were heavy curtains on the back wall, fabric wrapped acoustic panels on the sidewalls, and the floor was covered with low pile loop carpet. Other furnishings included metal-framed office chairs and equipment shelving. The other test room, the “Eargle Theater”, at 2.7 m. (H) x 5.3 m. (W) x 7 m. (L), was representative of a high-end, purpose built home theater. Acoustical treatment included cloth covered wall cavities with absorptive material, leather theater seating, heavy curtains, and some diffusive elements.

Using a custom-built manikin outfitted with DPA 4060 omni-directional microphones placed at the blocked canals, the Binaural Room Impulse Responses (BRIR’s) were measured for a range of head orientations at the listener's location in two different rooms. A 5 degree angular resolution was used.

Sensitivity to rotation depends highly on the audio material. In the case of pink noise, a source rotation of 2.5 degrees could make an audible difference, which was expected since earlier studies have shown that localization blur in the horizontal plane can be less than 2.5 degrees [5]. However, even a change in source direction of 10 degrees was found difficult to notice with certain music signals [6]. In fact, the angular resolution for the BRS scan was not directly comparable to localization blur as they are two different issues. The latter relates to a stationary head and moving source, the former relates to a “stationary” source and moving head. In work by Welti and Zhang [7], it was found that listeners had difficulty discriminating between 5, 10 and 15 degree angular resolution for BRS playback of music signals. For these reasons, a horizontal resolution of 5 degrees was deemed to be acceptable, especially in light of the fact that source localization was not a parameter under test. Binaural impulse responses were taken in 5 degree increments from -40 to +40 degrees. Previous studies have shown that head tracking should not be restricted to under +/- 30 degrees from front, in order to achieve a sense of naturalness [8].

During playback, program material was convolved with the BRIR filter sets using a Matlab based real-time convolution engine. The headtracker monitored the angular position of the listener’s head and sent the updated coordinates to the BRS playback software that switched to the appropriate BRIR filters corresponding to that angle. A correction for the headphone calibration was also applied as described below.

A diagram of the BRS measurements and playback systems are shown in Appendix 1.

2.3 Calibration of Playback Headphones

For the BRS tests, a Sennheiser HD518 circumaural headphone was chosen as the reference headphone. There were several criteria that made it a good choice for this task: its relatively wide bandwidth, smooth frequency response (not flat, but smooth), and consistent fit and seal on the listeners’ head and ears necessary for repeatable sound reproduction.

The Sennheiser HD518 headphone was measured on a GRAS 45AC test fixture equipped with an IEC 711 coupler. Figure 1 shows the measured frequency response of the left and right channels of the Sennheiser HD518 (solid curves). The curves were based on an average of five reseats of the headphone with the goal of minimizing errors related to headphone positioning on the ear simulator.

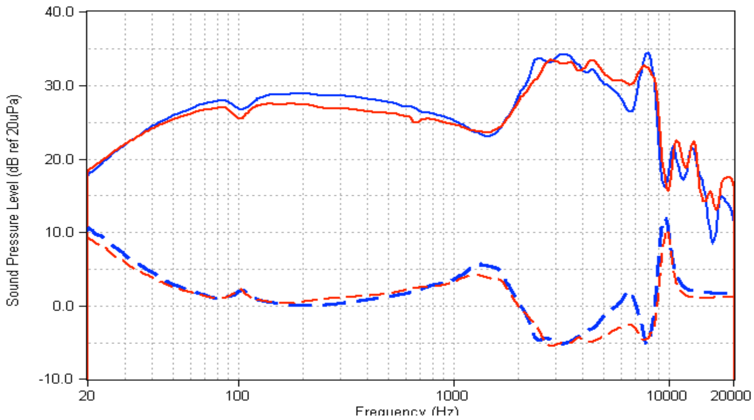


Figure 1 The left and right channel frequency responses of the Sennheiser HD518 headphones (solid curves) and the equalization (dotted curves) applied to flatten the response of the headphones.

The dashed curves in Figure 1 show the transfer function of the left and right channel equalization filters applied to “flatten” the reference headphones prior to application of the in-room target equalization. Schärer and Lindau observed that headphone equalization of the binaural signals is essential for natural sound color [9]. The headphones were equalized over a range of 25 Hz to 10 kHz. Equalization above 10 kHz can be problematic due to the increasingly variable response of the headphones. Figure 2 shows the response of the headphone before (dotted curves) and after (solid curves) flattening. This flat baseline response was used in conjunction with each of the 4 in-room target response curves, thus ensuring that any audible differences among curves were likely related to differences in their frequency responses, and not the response of the reference headphone. Finally, note that this calibration was based on a dummy head measurement, not an individualized listener calibration.

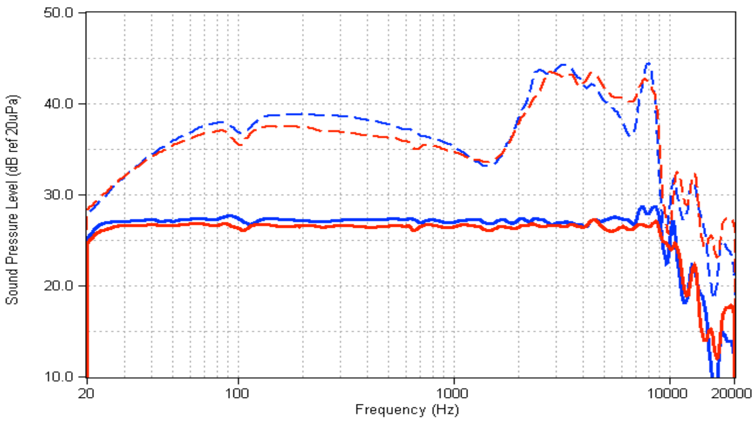


Figure 2 The dotted curves show the frequency response of the left and right channels of the headphone. The solid lines show the same after equalization.

2.4 Flattening of In Situ Loudspeaker Responses and Creation of In-room Target Response Curves

In the BRS validation test, subjects were asked to rate their preference amongst four different in-room target response curves applied to two different loudspeakers systems in two different rooms, the Living Room and the Eargle Theater, both of which are demonstration rooms at Harman International’s Northridge campus. For the in situ tests, the loudspeakers were individually measured using the Harman Audio Test System (HATS) with a nine-microphone array centered around the listening position on a 47.5 cm (18 in) grid. A spatial average of these nine positions was then equalized for each loudspeaker in order to create a flat response. Additional equalization was then applied to obtain the different in-room target response curves. Three of the in-room target curves, A, C and D, are shown in Figures 3(a) – (c) The fourth "curve", B, utilized no equalization thus leaving a flat room response.

The in-room target responses applied to both the in-situ and BRS playback systems were designed using the auto-EQ function in HATS, and implemented using the BSS Audio BLU-800 and BLU-80 DSP, which utilizes IIR filters.

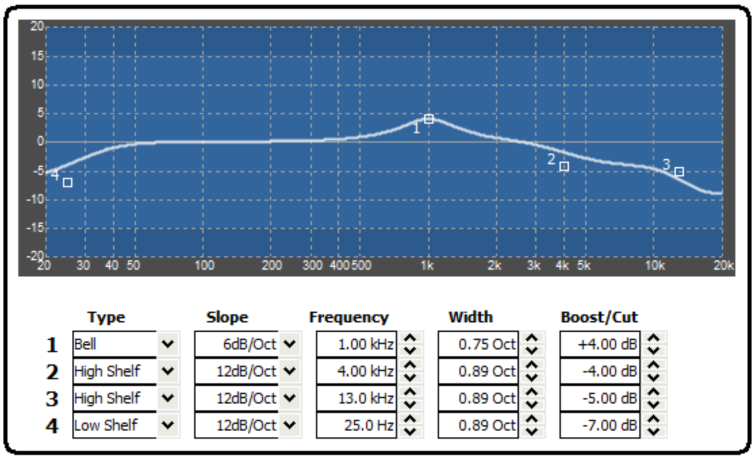


Figure 3a - Target room response curve 'A'

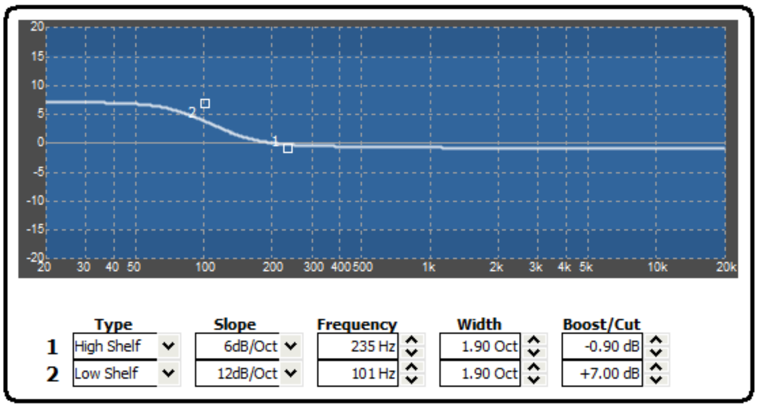


Figure 3b - Target room response curve 'C'

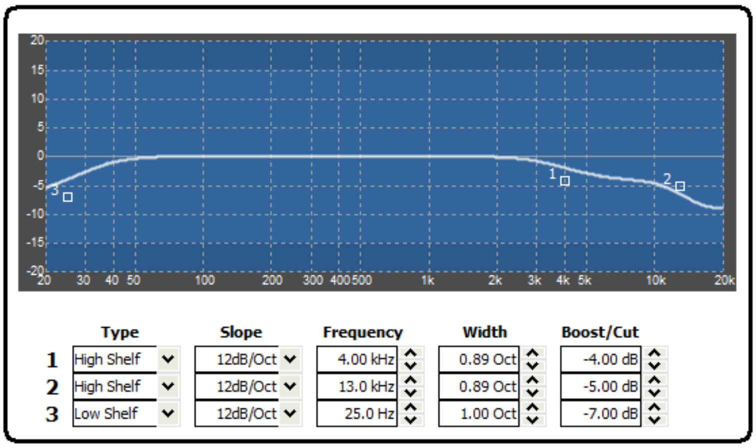


Figure 3c - Target room response curve 'D'

2.5 Program Material

Four different 5.1 music programs were selected and edited into short 20-30 second loops. They included an orchestral film score, Steely Dan's "Gaslighting Abbie", Blue Man Group's "Sing Along" and Toy Matinee's "Last Plane Out". All four programs have broad and smooth long term spectra, and have been found to be revealing of spectral features based on previous tests and listener training exercises.

Appendix 2 shows spectral analysis of the program material. For each program loop long term power spectral density (Welch method) was calculated for each channel, and the results summed to represent overall content.

2.6 Selection of Listeners

A total of nine listeners: three females and six males, were selected for the tests, all of whom were Harman employees paid for their participation. The listeners had audiometric normal hearing, and passed a listener training task that required them to reliably identify and discriminate among different spectral distortions added to different music programs. The listeners' experience with formal listening tests ranged from 20 years to a few months.

2.7 Test Design

The validity of the BRS system was tested using a 2 x 4 x 4 x 4 x 2 repeated measures ANOVA design, with the following independent variables: two test methods (BRS, in situ), four room response curves (A through D), four programs (film score and three popular music pieces), two rooms, and two observations. The dependent variable was preference rating. For the listening tests, the in-room target curves were comparatively rated on an 11-point preference scale that had semantic differentials on every second interval labeled: Really Dislike (1), Dislike (3), Neither Like/Dislike (5), Like (7), and Really Like (9). A strong preference between target curves was indicated by a separation in preference ratings of ≥ 2 points, a moderate preference ≥1 points, and a slight preference ≤ 0.5 rating.

In order to directly compare and measure the effects of the two different test methods as accurately as possible, other aspects of the experimental design and listening test conditions were held constant. For both BRS and in situ methods, the stimuli were presented the same number of times, at the same playback level.

A custom software application known simply as Harman Listening Test Software (LTS) was used to administer the experiments. This included control and switching of sound files, control of the BSS audio DSP, as well as the collection and storage of the listener response data. The software automatically checked for tied ratings and prompted the listener to make a forced choice between the test objects. Each listener participated in two listening sessions conducted on different days lasting approximately 20-30 minutes each. In one session, the listener evaluated the room response curve using the BRS playback system, and in another session, they evaluated the room response curve in situ. The order of the test sessions was randomly assigned to each subject in a balanced way, so that any learning or order effects were equally distributed between the two test methods. In each test session the listener completed 8 trials (4 programs x 2 observations). In each trial, the listener made comparative judgments between the four in-room target response curves until their final preference ratings were recorded. The listening test software application then automatically loaded the next trial, randomly reassigning a letter (A through D) to each room response curve. This test procedure was administered for each of the two listening rooms in random order.

2.8 Test Setup and Hardware

The hardware for the BRS listening test setup consisted of a Windows laptop, a digital sound card (RME Fireface UC) a programmable digital signal processor (BSS Audio BLU-800 and 80) and ultrasonic headtracker (Logitech). The hardware used for the Living Room in-situ listening test consisted of the same laptop, digital sound card and programmable digital signal processor in addition to a Harman Kardon 7550HD AV receiver / amplifier, and Harman Kardon HKTS 60 5.1 speaker system. The Eargle Theater in-situ listening test setup was similar to that of the Living Room with the difference of a three JBL Project Everest DD66000 left, center and right speakers, six JBL S4Ai THX Ultra2 surrounds and four JBL S1S-EX 18-inch THX Ultra2 subwoofers. The electronics in the system include a Mark Levinson No. 502 media console/surround processor, seven bridged JBL Synthesis S820 amplifiers (for the L, C, R speakers and subwoofers), and a JBL Synthesis 7 x 160-watt seven-channel amplifier (for the surrounds).

The Logitech headtracker utilized has an orientation resolution of 1/10 degrees and a maximum latency of 30 milliseconds. The latency of the entire BRS playback system is around 50ms, which does not cause audible effects under normal listening conditions. Work by Brungart, Kordik and Simpson has demonstrated that a tracking latency of at most 60ms is required to perceive a stable sonic image [10].

The relative playback levels of the different target response curves and program material were adjusted for equal loudness based on the ITU-R1770.2 loudness standard and verified by informal listening to assure that there were no discernible loudness differences between test curves. The average playback level of the music was adjusted to a comfortable level of 80dBA for both in-situ and BRS listening tests.

For the in situ portion of the experiments, listeners were positioned centrally in a standard 5.1 setup based on ITU-R BS.1116 listening room recommendations (matching the position of the dummy head used for scanning). The background noise levels were measured to be 25dBA in the Eargle Theater and 34dBA in the Living Room.

The BRS portion of the experiments were conducted with the listeners sitting in an office cubical, where the background noise level was measured to be between 35 and 40dBA. A key feature of conducting the BRS listening tests in a cubical is that it provided no visual reference to the two scanned rooms.

3 RESULTS

Section 3 reports the statistical analysis and results of the listening test.

3.1 Statistical Analysis

The data was analyzed using a repeated measures analysis of variance (ANOVA). A complete factorial analysis was used in the ANOVA model with a significance level of 0.05 utilized for all statistical tests.

3.2 Results Based on Method

Figure 4 shows combined data for the in-situ and BRS playback conditions for all programs and both rooms, with 95% confidence intervals shown.

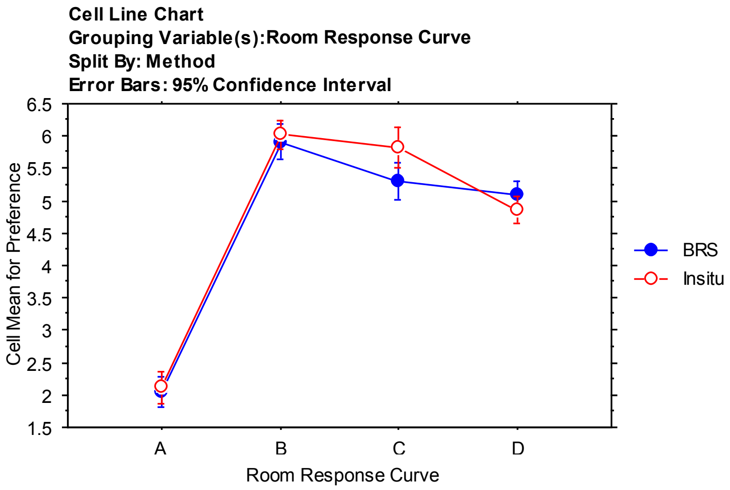


Figure 4. Combined data for the two playback conditions for all programs and both rooms. Data includes all subjects.

Reviewing the data for individual listeners, there were two listeners who had different ratings for repeated observations of the same stimuli. Figure 5 shows averaged ratings for observations O1 and O2 for each listener. Ideally a subject should repeat the same score for repeated ratings of the same stimuli. It can be seen in Figure 5 that subjects 363 and 379 were significantly less consistent in this regard. These subjects were two of the least experienced listeners, which may have resulted in less consistent ratings. It is also possible that the non individualized headphone calibration used in the BRS playback caused more errors for these particular subjects. These subjects were therefore removed from subsequent analysis. The result is shown in Figure 6.

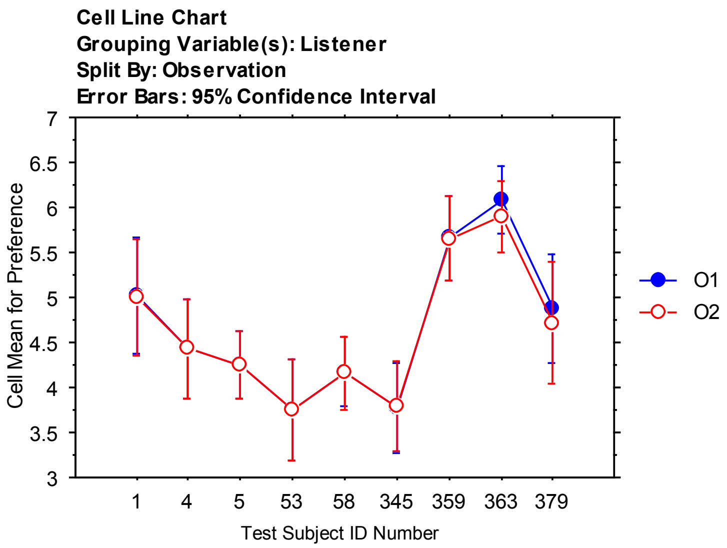


Figure 5. Repeatability of test subjects for repeated observations of the same stimuli.

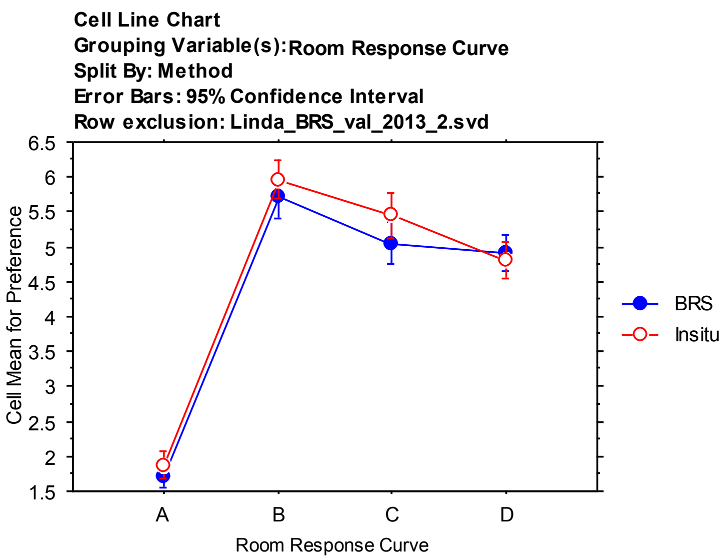


Figure 6. Combined data for the two playback conditions for all programs and both rooms with two subjects removed.

To determine if there was a statistical difference between ratings for the in-situ and BRS playback conditions, data from the listening tests was checked for Gaussian distribution and subjected to a repeated measures ANOVA test. The following were observed:

"Room Response Curve" was significant:
$F(3,24) = 16.47$, $p < .0001$.

"Observation" was not significant:
$F(1,8) = .49$, $p = .510$.

"Program" was not significant:
$F(3,24) = 1.08$, $p = .384$.

"Room" was not significant:
F (3,24) = 0.01, p = .920.

"Room*Room Response Curve" was significant:
F (3,24) = 3.83, p = .028.

"Method" was not significant:
F (1,8) = 1.34, p = .291.

4 DISCUSSION

The first three statistical observations in the previous section are more or less expected, but do help to confirm the validity of the BRS test method. The room response curve should have a large influence on ratings (or the test would be ill-designed). Since the order of the trials was randomized at all levels, a statistical difference across repeated observations would not be expected, unless there was a learning effect. Program should not be a factor, since subjects were asked to judge preference of the playback quality and not the program. However, program can be factor if it is not carefully selected, which was not the case here. Furthermore, in these types of tests, it sometimes occurs that there is an interaction between room response curve and program, but again, this was not the case here.

Scheffe for Preference
Effect: Room Response Curve
Significance Level: 5 %

	Mean Diff.	Crit. Diff.	P-Value	
A, B	-3.878	.359	<.0001	S
A, C	-3.472	.359	<.0001	S
A, D	-2.888	.359	<.0001	S
B, C	.406	.359	.0189	S
B, D	.991	.359	<.0001	S
C, D	.585	.359	.0001	S

Figure 7. Post hoc test for room response curve preference.

To verify that there was a perceived difference between the room response curves, a Scheffe post hoc test was completed to determine which pairs of response curves were statistically significant based on preference. With a significance level of 5%, Figure 7 shows that there was a statistically significant difference in preference amongst all pairs of room response curves.

As for the Room factor, there was no direct effect. The Room*Room Response Curve interaction was significant, however. This is interesting but not totally unexpected. This simply means that subjects had a different room response curve preference depending on which room they were listening in. Figure 8 shows this. It can be seen that there was a marked difference in preference for the 'D' room curve between the two rooms, for both playback methods. Since the sound systems in each room were equalized to the same measured target curves at the listeners seat, the difference in preference ratings is likely attributable to differences in directivity in the loudspeakers and/or acoustical properties of the rooms. The Eargle Theater had JBL Project Everest speakers, which had waveguides for mid/high frequency drivers and thus controlled directivity over a larger bandwidth, and the room itself is well damped acoustically. The Living Room had Harman Kardon HKTS 60 consumer speakers with small direct radiators and no directivity control. This room was smaller and with less controlled acoustics (curtains, carpet). It would not be surprising if these differences caused an interaction with the preferred room equalization curve. It is possible that differences in perception of direct and reflected sound had an effect on perceived spectral balance that was not reflected in the steady state measurements.

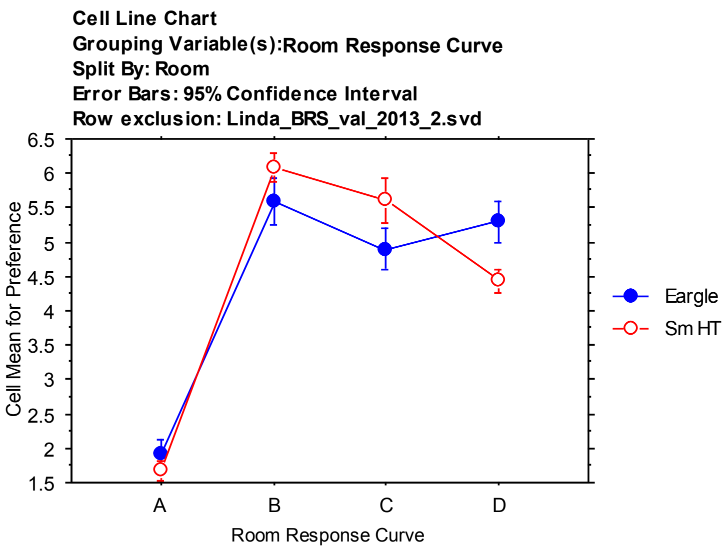


Figure 8. Preference for combined BRS data, for the two different listening rooms.

Finally, and most importantly, there was no statistical effect of "Method" on preference. Considered together, the results of the statistical analysis support the validation of BRS as an alternative to in-situ listening tests. It should be noted that you cannot prove the null hypothesis to be true simply because there is no evidence to prove it false. The lack of an effect due to the BRS method in this study doesn't preclude there will not be an effect if the test is repeated using a different set of experimental conditions or listeners.

Nonetheless, these results are encouraging and indicate that the BRS system has sufficient accuracy to allow listeners to discern differences in room response curves to the point where they can make repeatable preference ratings that are virtually the same as those made in situ.

Of interest to note is that these test results, like others before [2], were created using a generalized BRS system that was not calibrated to any particular listener participating in the experiment. The results show a high level of listener reliability between the two methods in absence of a more complex or individually calibrated system. The "robustness" of the BRS system used in these experiments makes it a viable option for listening tests in lieu of impractical in situ listening tests.

5 CONCLUSION

This paper reports the results of a validation experiment performed using a BRS system that captures and stores binaural room impulse responses of sound sources within a given listening space and then reproduces them via a headtracking headphone system.

The experiments investigated whether a listener's room response curve preference ratings made in situ were similar to ratings made using a BRS reproduction of the same room. The main conclusions of these experiments are as follows:

1) There were no significant differences in preference between the BRS and In situ Methods .

2) The main effect was due to the room response curve.

3) There was an interaction between the different listening rooms, and the room response curves. This interaction was confined to curve 'D'.

The extent to which these experimental results can be applied to other listening spaces is a question currently under investigation.

6 FUTURE RESEARCH

The research encompassed in this paper was centered around two home-theater type settings. Future research will be aimed at validating the BRS system for larger, cinema-sized spaces. In addition, spectral balance preferences between the two sets of experiments will be gathered and compared to see if listeners' perceptions of the sound system's spectral balance change with venue size. Future research will also focus on what the preferred in-room target response should be and how it's perceived sound quality is influenced by program and venue size.

7 ACKNOWLEDGEMENTS

The authors would like to thank Drs. Sean Olive and Trevor Cox for their support of this research and the sponsors of this research, Harman International. Special thanks to Elisabeth McMullin of Harman Corporate Technology Group for her assistance in setting up these experiments.

8 REFERENCES

1. Mackensen, P., Felderhof, U., Theile, G., Horbach, U., and Pellegrini, J., “Binaural Room Scanning—a New Tool for Acoustic and Psychoacoustic Research” J. Acoust. Soc. Am. Volume 105. (Issue 2) pp. 1343-1344 (1999).

2. Olive, S., Welti, T., and Martens, W., “Listener Loudspeaker Preference Ratings Obtained In Situ Match Those Obtained via a Binaural Room Scanning Measurement and Playback System”. Proceedings of the 122nd International Convention of the Audio Engineering Society, Preprint 7034 (2007)

3. Olive, S., and Welti, T., “Validation of A Binaural Car Scanning System for Subjective Evaluation of Automotive Audio Systems”, AES 36th International Conference, Dearborn, Michigan, USA(2009)

4. Postel, F.; Hegarty, P., and Bech, S., “A listening test system for automotive audio: PART 5 - The influence of listening environment on the realism of binaural reproduction”. Proceedings of the 130th Convention of the Audio Engineering Society, London, UK. (2011)

5. Blauert J.: Spatial Hearing. The Psychophysics of Human Sound Localization. Revised ed. 1997 (The MIT Press, Cambridge, Massachusetts, London, England.)

6. Hiekkanen, T., Makivirta, A. and Karjalainen, M.. "Virtualized Listening Tests For Louspeakers". J. Audio Eng. Soc., vol. 57, (2009)

7. Welti, T., and Zhang, X., “Angular Resolution Requirements for Binaural Room Scanning”, 129th Convention of the Audio Eng. Soc., preprint 8271, (Nov, 2010).

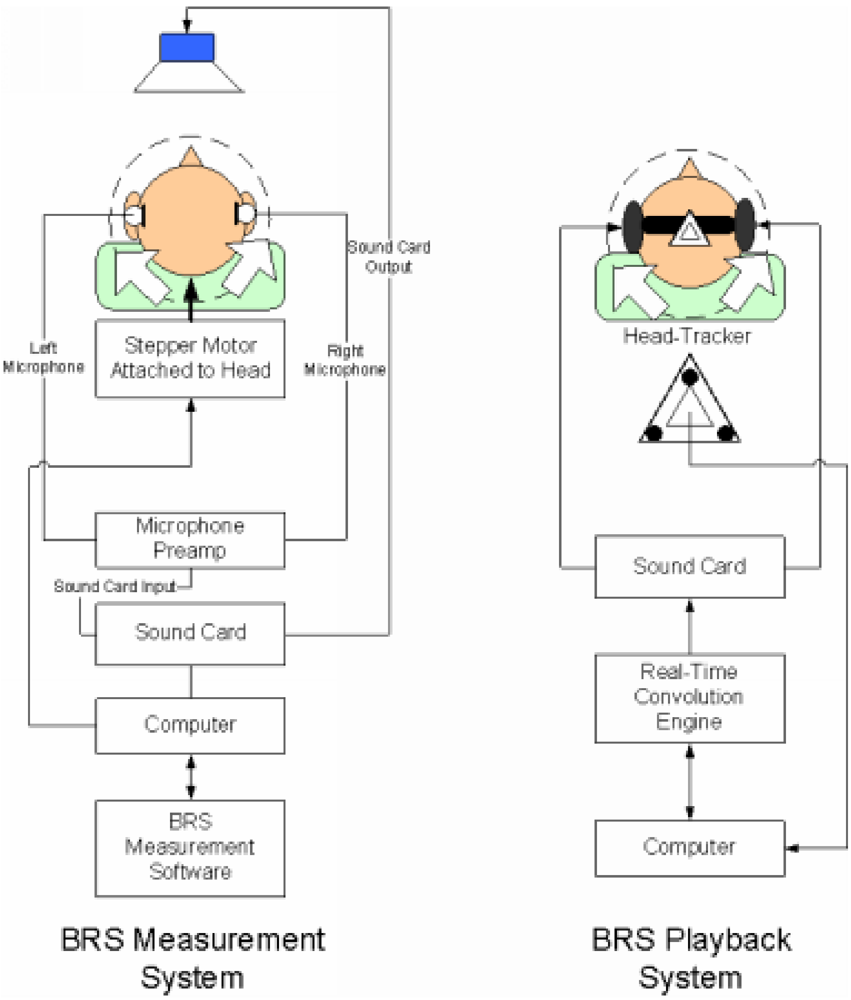
8. Laitinen, M. V., Pihlajamaki, T., Losler, S. and Ville Pulkki, V. "Influence Of Resolution Of Head Tracking In Synthesis Of Binaural Audio". Proceedings of the Audio Engineering Society 132nd Convention. Budapest, Hungary. (2012).

9. Schärer, Z., and Lindau, A. “Evaluation of Equalization Methods for Binaural Signals”. Proceedings of the Audio Engineering Society 126th Convention. Munich, Germany. (2009)

10. Brungart, D.S.; Kordik A.J.; and Simpson, B.D.: Effects of Headtracker Latency in Virtual Audio Displays. J. Acoust. Soc. Am., Vol. 54, No. 1/2, (2006)

APPENDIX 1

Below are block diagram of the binaural room scanning (BRS) measurement and playback system.



APPENDIX 2

Long-term spectral power density of the four pieces of program material. Levels normalized at 500Hz.