Audio Engineering Society

Convention Paper 7196

Presented at the 123rd Convention
2007 October 5–8 New York, NY, USA

The papers at this Convention have been selected on the basis of a submitted abstract and extended precis that have been peer reviewed by at least two qualified anonymous reviewers. This convention paper has been reproduced from the author's advance manuscript, without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA; also see www.aes.org. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Interaction between Loudspeakers and Room Acoustics Influences Loudspeaker Preferences in Multichannel Audio Reproduction

Sean E. Olive[1] and William L. Martens[2]

[1] Harman International Industries Inc., Northridge, California, 91329, USA
solive@harman.com

[2] Schulich School of Music, McGill University, Montreal, Quebec, H3A 1E3, Canada
wlm@music.mcgill.ca

ABSTRACT

The physical interaction between loudspeakers and the acoustics of the room in which they are positioned has been well established; however, the influence on listener preferences for loudspeakers that results from such variation in room acoustics has received little experimental verification. If listeners adapt to listening room acoustics relatively quickly, then room acoustic variation should not significantly influence loudspeaker preferences. In the current study, two groups of listeners were given differential exposure to listening room acoustics via a binaural room scanning (BRS) measurement and playback system. Although no significant difference in loudspeaker preference was found between these two groups of listeners, the room acoustic variation to which they were exposed did significantly influence loudspeaker preferences.

1. INTRODUCTION

There are many scientific studies showing the physical effects of acoustic interactions between the loudspeaker and listening room that are well summarized in a recent paper by Toole [1]. Below a transition region around 300 Hz, the room adds its own set of resonances that cause large seat-to-seat variations in the frequency response that change depending on the locations of the loudspeakers. Above 300 Hz, the quality and proportion of the direct and reflected sounds at the listener are influenced by the acoustical properties of the loudspeakers and room, and the locations of the listeners and loudspeakers. The physical effects of these interactions are easily seen in acoustic measurements made at the listening locations. However, the perception of these loudspeaker-room interactions and their effect on the quality of recorded and reproduced sound are generally not well understood.

Some studies have reported that that the position of the loudspeaker can produce preference differences that, in some cases, are larger than those measured among different models and brands of loudspeakers [2]-[8]. Few studies have shown how listeners' loudspeaker preferences vary when the loudspeakers are compared in different listening rooms [3]-[6]. One study that did, found that four different listening rooms had no significant effect on listeners' loudspeaker preferences [4]. This was not the expected or desired result, since the researchers were developing a room-adaptive loudspeaker aimed at correcting loudspeaker-interactions [5]. There was an expressed fear that room-adaptive speakers might fall under the same category as exotic audio cables:

"If it cannot be demonstrated in a convincing fashion that different rooms affect the sound quality of loudspeakers, then it will be difficult to sell a product whose purpose is to correct what cannot essentially be heard." (p. 67)

To extricate themselves, the researchers did something quite unnatural that would exaggerate the effect of the room on loudspeaker preferences. Through the means of a binaural recording-playback system, listeners were able to make immediate comparative ratings of the same loudspeaker among the four rooms with no time gap in between the comparisons. This produced the opposite but desired result. Now the listening rooms became the dominant factor in listeners' preferences, and the loudspeaker variable had no effect. The researchers were quite happy with the large room effect because it meant funding for the room-adaptive loudspeaker project was no longer in jeopardy.

The above experiment had two important findings: 1) listeners were much better at adapting to the room acoustics than were the room-adaptive loudspeakers, and 2) the experimental results could be manipulated by changing the order and context in which the stimuli were compared. Listeners either adapted 100% to the room acoustics, or not at all, depending on the context under which the rooms and loudspeakers were compared.

Room acoustic adaptation is defined as a change in the listener's response to the room acoustics after having spent some period of time in the room. In the context of subjective evaluation of different loudspeakers in different rooms, room acoustic adaptation would be observed as a change in how the room affects the loudspeaker ratings Within certain limits, more time spent in the room should produce more adaptation (less room effect), whereas less time spent in the room should produce less adaptation (greater room effect). However, the time course of adaptation to room acoustics as an influence on the perception of reproduced sound is largely unknown. Therefore, a manipulation that is intended to modulate room acoustic adaptation may or may not work.

The precedence effect [9]-[29] and spectral compensation [30]-[32] are two well-known room acoustic adaptation mechanisms that permit humans to perceive the timbre, direction and intelligibility of a sound source in reflective spaces. Both mechanisms involve a central auditory cognitive decision process that uses current and prior learned auditory information, and cues from other sensory modalities. Information that is considered redundant, irrelevant or implausible is suppressed, indicating adaptation has occurred [17].

Research in room acoustic adaptation has important ramifications on the methods used for conducting listening tests, and the design of loudspeakers and rooms used for audio recording and reproduction. The current standards for designing listening rooms [33], [34], loudspeakers and loudspeaker-room-interaction correction products may not adequately account for human's perception and adaptation to room acoustics. The application of technology to solve loudspeaker-room problems may be superfluous if human perception has already taken care of them.

However, there are many unanswered questions about room acoustic adaptation. The parameters or listening conditions, under which room adaptation works well, or not at all, are not well established. Room acoustic adaptation has not been studied under multichannel audio listening conditions, where the precedence effect and spectral compensation may also operate differently than expected. The room acoustics, number of loudspeakers, the complexity of the listener's task and the attention it requires, the test signals and recording techniques may all influence how well these mechanisms work and how much adaptation occurs.

There are many methodological challenges in room acoustic adaptation research. The independent variables (i.e. different loudspeakers, rooms, loudspeaker positions, room exposure time) must be manipulated in a way that permits real-time, double-blind, comparative evaluations. This is not possible or practical using in situ listening test methods. For this reason, a binaural room scanning (BRS) method was chosen for this study because it allows the independent variables to be captured and stored as a set of binaural room impulse responses (BRIR), that may later be reproduced through high quality headphones equipped with a low-latency head-tracking system [35]-[37]. Details on the BRS system used in this study are in [36], including a description of its calibration and performance. A validation test was conducted that determined the BRS method produced virtually the same loudspeaker preference ratings as measured using in situ loudspeaker listening tests

The following sections describe an experiment designed to measure the influence of loudspeaker and room acoustic interactions on listener preference for multichannel music imagery, and the role room acoustic adaptation plays in changing these preferences. Section 2 defines the main research questions and the experimental design and methods. Section 3 describes the experiment results, with a discussion of their meaning in Section 4. The main conclusions and limitations of the experiment are defined in Section 5.

2. EXPERIMENT

2.1. Research Questions

An experiment was designed to address the following research questions:

(1) To what extent are listeners' preferences for multichannel music imagery influenced by different loudspeakers, room acoustics and interactions between the two variables?

(2) Can the effects and interactions on listener preference be explained by acoustic measurements of the loudspeakers, room acoustics, and their interaction?

(3) To what extent does room acoustic adaptation diminish the effect of room acoustics on listeners' preference ratings?

2.2. Experiment Design

To test these research questions an experiment was designed to measure listener preferences for multichannel music imagery reproduced through four different loudspeakers, located in four different listening rooms. The program selections consisted of three different 5-channel music selections. The stimuli were evaluated by two separate groups of listeners using two different trial ordering schemes that provided a differential exposure to the room acoustics, and hence, a potential change in room acoustic adaptation. The independent variables and their levels are summarized in Table 1. The experimental design was a 2 x 4 x 4 x 3 x 3 repeated measures analysis of variance (ANOVA), where the between-subjects factor was the trial ordering method. More details on each independent variable are given in the following sections.

Table 1: The experiments independent variables and their levels

Independent Variable	Levels of Variable
Trial Ordering Method	Successive, Intermixed
Loudspeaker	L1, L2, L3, L4
Room	R1, R2, R3, R4
Program	JV, LL, SD
Observation	O1,O2,O3

2.3. Independent Variables

2.3.1. Listening Rooms

Four rectangular listening rooms were used in these experiments, all located at Harman International in Northridge, California. The rooms are used for both formal and informal subjective evaluations of audio technology, and are intended to acoustically mimic typical domestic listening spaces. All four rooms have sufficiently low background noise (<25 NC), and fall within typical domestic room sizes, ranging in volume from 60 to 151 cubic meters [38]. Dimensional diagrams for the four rooms and their loudspeaker setups are shown in Appendix 1.

Apart from their dimensional differences, three of four rooms (R1, R2 and R3) had similar acoustic treatment in terms of the amount, type and distribution of absorption and scattering objects about the room. The acoustic treatment was mostly provided by common objects found in most households such as carpeting, curtains, chairs and bookshelves.

In contrast to the other three rooms, Room R4 was more reflective. This was achieved by removing most of its furniture (6 leather tub chairs) and several 1.22 m (H) x 1.22 m (W) X 7.6 cm (D) fiberglass panels from its two side walls. After doing so, the two side walls were essentially 100% reflective. The hardware floor was also reflective except for a 2.7 cm woolen rug with foam underlay covering 30% of the floor under the seating area. The only significant absorption in the room was provided by some fiberglass panels that covered 42% of the area of the front and rear walls.

Table 2: Below are the listening room dimensions and volumes.

Room	Dimensions (m) (L x W x H)	Volume (m^3)
R1	5.84 x 4.29 x 2.4	60.19
R2	6.2 x 5.08 x 2.74	105.55
R3	9.04 x 6.58 x 2.54	151.2
R4	7.26 x 6.32 x 2.74	126.03

Some limited diffusion above 1 kHz was provided by RPG Skylines [39] mounted near the ceiling along the four walls. A feature common to all rooms was a reflective surface where the first lateral reflections occurred at the listening location. The properties of these reflections varied among the different rooms based on the positions of the listener and loudspeakers from the side walls. In the two smallest rooms (R1 and R2), the first lateral reflections were approximately the same. In the largest room (R3) [40], the lateral reflections were much later and lower in level due to the greater distances between the loudspeakers and the side walls.

2.3.2. Loudspeaker Selection

There were several criteria for selecting the loudspeakers in these experiments. First, there was a desire to have constant on-axis and low frequency performance among all four speakers, with differences only in their measured off-axis performance. By controlling the direct sound, loudspeaker preferences could be directly related to the physical properties of the loudspeaker's off-axis radiated sound and the reflective properties of the listening rooms. Furthermore, the acoustic similarities among the loudspeakers would minimize stimulus recognition biases that might otherwise swamp any room effects, and possibly inhibit room acoustic adaptation.

To improve the external validity of the study only loudspeaker designs most commonly found in consumer and professional audio setups were selected. A custom "adjustable" loudspeaker was designed and fabricated using forward-facing, direct-radiator, electro-dynamic components, which make up the majority of all loudspeakers sold. On each of its four sides, an array of drivers were configured to simulate a common 3-way bookshelf or tower loudspeaker, with and without a waveguide on the tweeter, and two different center-channel 3-ways that had two midrange drivers arranged either vertically and horizontally on either side of the tweeter. Details on these four loudspeakers are given in Table 3. The loudspeaker had an active, programmable cross-over processor (DBX DriveRack 260) that provided driver equalization. A subwoofer (JBL HTPS400) was common to all four loudspeakers with the arrays driven by a 3-channel amplifier (Proceed Amp 3).

The anechoic frequency response measurements (see Appendix 2) show that each loudspeaker has virtually identical flat (± 1 dB, 30 Hz-20 kHz) on-axis frequency response. The only differences occur in their off-axis responses above 300 Hz, as indicated in their spatially-averaged early-reflection and sound power curves, and their directivity indices. The off-axis curves of the four loudspeakers get progressively worse in the following order: L2 (best), L1, L3 and L4 (worst). It is arguable whether L4 is worse than L3, but it is safe to say that L3 and L4 are technically not as good off-axis as L2 or L1. A photograph of the speaker and its associated equipment is shown in Fig. 1.

Table 3: A description of the four loudspeakers used in these experiments.

Loudspeaker	Configuration	Cross-over Frequencies
L1	3-way, with 25 mm tweeter without waveguide, 165 mm midrange, 203 mm woofer	2 kHz, 150 Hz, 80 Hz
L2	3-way, with 25 mm tweeter with waveguide, 165 mm midrange, 203 mm woofer	Same as above
L3	3-way vertical array with 25 mm tweeter, 2 x 102 mm mid-range	3 kHz, 150 Hz, 80 Hz
L4	3-way horizontal array with 25 mm tweeter, 2 x 102 mm mid-range	Same as above

Figure 1 A photograph of the adjustable loudspeaker with its electronics, showing the L2 configuration. Note that the top midrange driver is not active for this configuration.

2.3.3. Program Selection

Three music selections were used in these experiments, all originating from commercially available multichannel DVD-A and DVD discs (see table 4).The music tracks were transferred from DVD to a digital editing software program (Sonar 6, Producer) and edited into short 20-30 s loops. Each loop was saved as 5-channel (16 bit, 48 kHz) Microsoft extensible wav file. All three tracks were well-recorded, with full-range, spectrally-dense, auditory imagery distributed across all five channels. This ensured all programs would be as equally revealing and sensitive to possible spectral and spatial differences related to the different loudspeaker-room combinations.

Table 4: Program selections used in the experiments

Program	Source
JV (Jazz Vocal)	Gordon Goodwin's Phat Pack, "The Phat Pack", CD-DVD (2006), Track 2: Too Close for Comfort (featuring Dianna Reeves)
LL (Lyle Lovett)	Lyle Lovett, Joshua Judges Ruth DVD Audio/DTS (2002), Track 2: Church
SD (Steely Dan)	Steely Dan, Two Against Nature, DVD Audio (2001), Track 1: Gaslighting Abbie

2.3.4. Successive and Intermixed Treatments of Room

The experimental context in which auditory stimuli are presented has been shown to influence preference choices or ratings [41]. The context under which different loudspeakers [42], [43] and listening rooms [4] are compared can affect how listeners scale their preferences. Similarly, trial ordering of music programs influences preference choices for different multichannel microphone techniques [44], [45]. Some researchers have argued that context influences listeners' preferences to such a degree that measures of preference should be avoided altogether in audio research [46]. This is a hasty conclusion since it ignores the fact there are methods for dealing with contextual effects in preference experiments, and that humans make preference choices every day where context is an integral factor in their decisions. To remove context and preference from sound quality listening tests severely limits their external validity and generalization to how humans respond in the real-world. To fully understand the complex cognitive decision processes behind listeners' loudspeaker preferences scientists should not exclude the influence of context in listening tests. For these reasons the authors view context, not as a nuisance variable, but as integral factor worthy of investigation [47].

To study how preference and room acoustic adaptation might be influenced by context, two different trial ordering schemes were used to evaluate the stimuli. The two trial ordering schemes referred to as successive and intermixed treatments [49] are graphically shown in Fig. 2. Note that P1-P3 represents the three different programs used in these experiments. The trial and session orders were randomized in a balanced way for both successive and intermixed treatments to control for any order-related biases.

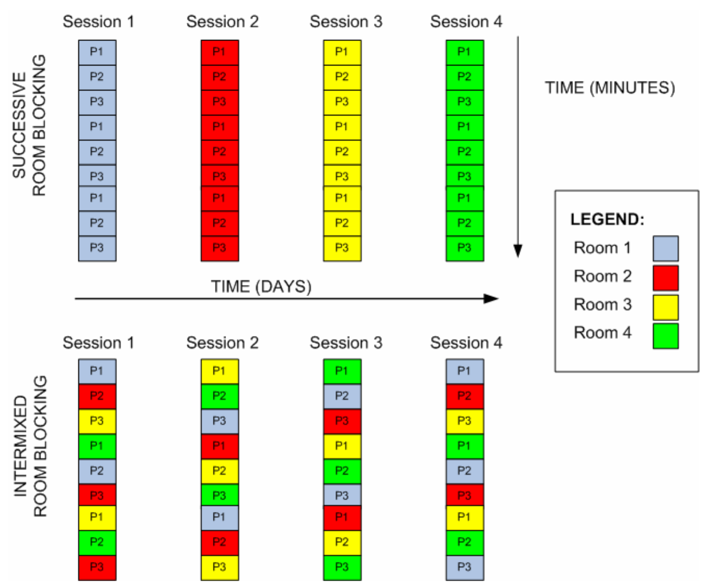


Figure 2 A graphical representation of successive (top row) and intermixed treatments (bottom row). Each column represents a different block or listening session. P1- P3 represent the three program selections. The order of program and session was randomized among subjects.

One group of listeners evaluated the stimuli using a successive treatment where room was held constant throughout a single listening session, and changed only between subsequent listening sessions. A second group of listeners used an intermixed treatment, where the room was always changed in subsequent trials within the listening session. The differential exposure to the room acoustics was expected to produce differences in room acoustic adaptation. Based on a similar study where successive treatments of room were used [4], it was believed that the successive treatment would produce more room acoustic adaptation (i.e. less room effect on the preference ratings) due to the longer exposure (nine trials) to the same room acoustics. The intermixed treatments were expected to produce less room acoustic adaptation (i.e. more room effect on the preference ratings) due to the shorter exposure time (one trial) in the same room.

2.3.5. Dependent Variable (Preference Rating/Choice)

All stimuli in these experiments were evaluated on an 11-point preference scale that included semantic descriptors on every second interval to indicate the intensities of the likes/dislikes for the scaled stimulus (see Fig. 2 of Appendix 3). Each listening session began with a block of 6 trials, where the listener made preference choices between paired comparisons of the four loudspeakers (see Fig. 1 of Appendix 3). The purpose of these pre-trials was to familiarize the listener to the stimuli using an easier task than a multiple preference rating task [50]. The paired comparison task provided the means to study how the complexity of the listener's task affects their performance and preference. The simpler task would take much less time to complete, and possibly produce less room acoustic adaptation than the multiple comparison preference rating task.

2.3.6. Selection of Listeners

A total of 23 listeners voluntarily participated in the listening experiments, all of whom were paid. The listeners were current or former employees of Harman International and were screened for normal audiometric hearing [50]-[51]. Each subject passed a training task requiring them to identify spectral distortions added to music [52]. Eight of the 23 listeners (35%) could be considered as experienced having 2 or more years of experience in controlled loudspeaker listening tests. The majority of listeners (65%) were relatively inexperienced listeners having less than 6 months of experience in formal loudspeaker listening tests.

The ages of the listeners ranged from 22 to 48 years (median age = 31, SD = 8.6 years), and 90% were male. This should not unduly impact how generalized the results are since previous studies by Toole [54]-[55] and Olive [56] have shown gender not to be factor in loudspeaker preference. All listeners were given written instructions (see Appendix 3) before beginning the experiments.

2.3.7. Loudspeaker Setup and BRS Measurements

Loudspeaker position, loudspeaker setup and listening position were treated as nuisance variables in this study. To control positional and setup biases, all loudspeakers were scanned in the same positions and set up symmetrically in all four rooms according to ITU-R BS.775-1 [33]. Consequently, there was no loudness, timbre or spatial imbalances between the left and right channels in any of the rooms. This makes it easier to account for and explain possible effects and interactions between the loudspeakers and rooms. A disadvantage of using only symmetrical setups and rectangular rooms is that the results of this study may not be well-generalized to loudspeaker setups more commonly found in consumer's homes.

In each of the four different listening rooms, BRIR measurements were taken at the listening location, for each loudspeaker, at each of the five positions. The distances and angular positioning of the loudspeakers relative to the listener were compliant with ITU-R BS.775-1 [33]:

- ± 30º for the front left and right loudspeakers
- 0º for the center loudspeaker, and
- ± 115º for left and right surround channels

The loudspeakers were symmetrically arranged in a circle so that each speaker was equidistant to the listening location in the middle of the circle, the so-called listening "sweet spot". Precise calibration of the loudspeaker's position in each room was performed through the use of a laser pointer attached to each loudspeaker, and a small white reflective dot on the binaural manikin's nose. Further confirmation of positional accuracy was done by inspecting the in-room frequency response and BRIR of each loudspeaker.

Precise loudspeaker positioning was necessary to make certain the direct sound was constant across all loudspeakers, positions and rooms in terms of its: (a) frequency response, (b) level, (c) relative time arrival among the five channels, and (d) angle of sound incidence at the listener location. The BRS measurement parameters are summarized in table 5.

Table 5: BRS measurement parameters for these experiments

Parameter	Value
Sampling Rate	48 kHz
Bit Rate	16 bits
Average in-room S/N	75 dB
Impulse Length (ms)	500 msec
Spatial Resolution	2° resolution ± 60° horizontal azimuth

2.3.8. BRS Playback

The BRS playback of the different loudspeakers and rooms was done using the calibrated playback system that is graphically shown in Fig. 3. The server computer (HP xw8200) controlled the entire listening test through its custom listening test software. The output from its digital sound card (M-audio Firewire 410) was sent to a master volume control (Lexicon MC-12 Balanced), after which it was sent to the 8-channel sound card (RME HDSP 965 with ADI-8 DS interface) connected to the BRS playback convolution engine. The output from the BRS engine was sent to a headphone amplifier (AKG phone amp V6HP) that powered the listener's headphones (Sennheiser HD 600). The azimuth of the listener's head position was monitored with a Logitech ultrasonic head-tracker that transmitted the current angle to the BRS playback engine. Upon receiving the current head position angle, the BRS engine switches to the corresponding set of BRIR filters.

The listener could switch at will between the different surround sound loudspeaker systems using a hand-held wireless Pocket PC (HP model hx2490) and a custom listener client software application known as "Pocket Car Evaluator". The Pocket PC sent control commands wirelessly to the server computer that would immediately switch to the BRIR sets of the specified loudspeaker surround system.

All listening tests were conducted in the Harman International reference listening room, one of the four listening rooms included in these tests. To enhance the realism and presence of the BRIR playback, a set of high quality loudspeakers were set up in the identical ITU-R configuration used for the BRIR measurements (see Fig. 4). The listener sat in the same sweet-spot where the binaural manikin was placed for the BRS measurements. The average playback level of the program selections at the listeners blocked meatus was a comfortable 76 dBA (slow). The playback level remained fixed for the duration of the experiment.

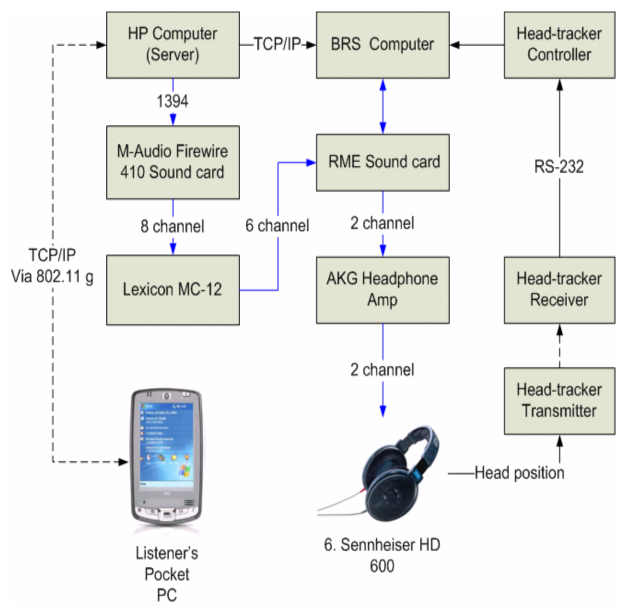


Figure 3 A block diagram of the equipment used for the listening tests and BRS playback.

2.3.9. Listener's Task

Each listening session began with a series of 6 pre-trials where the listener made a simple preference choice between different pairs of surround sound loudspeakers for one program selection (LL). The purpose of the pre-trials was to familiarize the listener with the stimuli using a less cognitively demanding task, before beginning the more demanding multiple-comparison preference rating task. Using the Pocket PC and the Pocket Car Evaluator software, listeners could switch at will between the stimuli and enter their responses (see Fig. 1 of Appendix 3). Once the final ratings were entered and stored in the database, the next trial would automatically begin.

Listeners generally completed all 6 preference choice pre-trials within 5-10 minutes and then begin their multiple-comparison preference rating task. Listeners were required to complete all 9 trials and enter their ratings using the GUI shown in Appendix 3 (see Fig. 2). The software automatically checked for tied ratings and forced them to make a preference choice between them. Typically the entire listening session took about 25-30 minutes to complete. Only two sessions were allowed on any given day, with a break between the morning and afternoon sessions.

Figure 4 A photograph of a listener in Room R4 (shown in its normal acoustic configuration) doing a listening test with the BRS playback system.

3. EXPERIMENT RESULTS

This section presents the statistical analysis of the results of the experiment described in the previous section. Only the results from the preference rating task are presented here; the results from the preference choice task will be the topic of a future paper.

3.1. Statistical Analysis

A repeated measures ANOVA for a 2 x 4 x 4 x 3 x 3 full factorial model was used to analyze the independent variables: Trial Order Method (2 levels), Loudspeakers (4 levels), Rooms (4 levels), Programs (3 levels) and Observation (3 levels). The dependent variable was preference rating. An alpha level of 0.05 was used for all statistical tests. The ANOVA Table in Fig. 5 indicates that Room was the only significant main effect; F (3, 60) = 4.14, p = 0.01. There were also significant interactions between the factors Room and Loudspeaker; F (9,180) = 3.11, p = 0.002, and the factors Loudspeaker and Program; F (6,126) = 2.71, p=0.02.

ANOVA Table for Preference Rating
Row exclusion: FinalALlListenersANOVA.svd

	DF	Sum of Squ...	Mean Square	F-Value	P-Value	Lambda	Power
Method	1	21.740	21.740	.197	.6617	.197	.070
Subject(Group)	21	2318.116	110.386				
Room	3	129.957	43.319	4.141	.0096	12.423	.836
Room * Method	3	37.334	12.445	1.190	.3209	3.569	.296
Room * Subject(Group)	63	659.065	10.461				
Loudspeaker	3	7.272	2.424	.201	.8952	.603	.085
Loudspeaker * Method	3	16.471	5.490	.456	.7143	1.367	.135
Loudspeaker * Subje...	63	759.213	12.051				
Program	2	24.846	12.423	1.952	.1547	3.904	.370
Program * Method	2	36.540	18.270	2.870	.0678	5.741	.523
Program * Subject(Gr...	42	267.324	6.365				
Observation	2	8.104	4.052	1.898	.1626	3.795	.360
Observation * Method	2	6.156	3.078	1.441	.2480	2.883	.281
Observation * Subject...	42	89.682	2.135				
Room * Loudspeaker	9	125.683	13.965	3.110	.0016	27.992	.978
Room * Loudspeaker ...	9	17.078	1.898	.423	.9219	3.804	.204
Room * Loudspeaker ...	189	848.611	4.490				
Room * Program	6	22.037	3.673	1.216	.3022	7.299	.458
Room * Program * M...	6	34.161	5.693	1.886	.0882	11.314	.680
Room * Program * Su...	126	380.435	3.019				
Room * Observation	6	18.141	3.023	1.255	.2831	7.529	.472
Room * Observation *...	6	8.273	1.379	.572	.7518	3.433	.220
Room * Observation *...	126	303.595	2.409				
Loudspeaker * Program	6	37.404	6.234	2.706	.0166	16.239	.860

Figure 5 The ANOVA Table for the preference rating test.

3.2. Main Effects: Room

The mean preference ratings and the 95% confidence intervals are shown in Fig. 5 for the factor Room. In order of decreasing preference, the mean ratings for each listening room were: R1 = 5.26, R2 = 5.16, R3 = 5.05 and R4 = 4.73.

A Scheffe post-hoc test was performed to determine if the differences in ratings between each pair of rooms were statistically significant (see Fig. 7). The results of this test indicate that rooms R1, R2 and R3 were all preferred over the Room R4. There were no significant preferences among these three rooms.

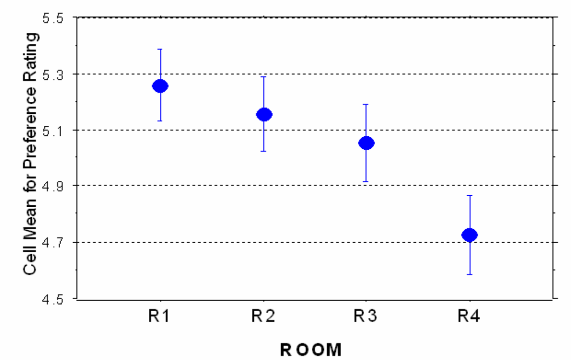


Figure 6 The mean preference ratings and 95% confidence intervals for the variable Room.

Scheffe for Preference Rating
Effect: Room
Significance Level: 5 %

	Mean Diff.	Crit. Diff.	P-Value	
R1, R2	.103	.269	.7651	
R1, R3	.206	.269	.2053	
R1, R4	.533	.269	<.0001	S
R2, R3	.103	.269	.7670	
R2, R4	.430	.269	.0002	S
R3, R4	.327	.269	.0093	S

Figure 7 The Scheffe post-hoc test Table for the variable Room.

3.3. Interaction Effects

3.3.1. Loudspeaker and Room Interaction

The ANOVA Table in Fig. 5 shows there was a significant interaction between the Room and Loudspeaker. The mean preference ratings and 95% confidence intervals for this interaction are plotted in Fig. 8 below.

The graph clearly shows loudspeaker preference ratings were dependent on the room in which the ratings were given. In Rooms R4 and R2 there were significant differences in loudspeaker preference ratings, yet in Rooms R1 and R3 there were none. Overall, the loudspeakers were rated the lowest in Room R4, and highest in Room R1. The loudspeakers with the largest variance in preference ratings due to the influence of Room were L2 and L4. The preference ratings for Loudspeaker L1, and to a lesser extent L3, were the least influenced by variations in room acoustics.

3.3.2. Program and Loudspeaker Interaction

The ANOVA Table indicates a second significant interaction between the factors Program and Room. This interaction, plotted in Fig. 9, indicates that program JV (jazz band with female vocalist) and Loudspeaker L4 are largely responsible for this interaction effect. When Loudspeaker L4 was auditioned using program JV, it received significantly lower ratings.

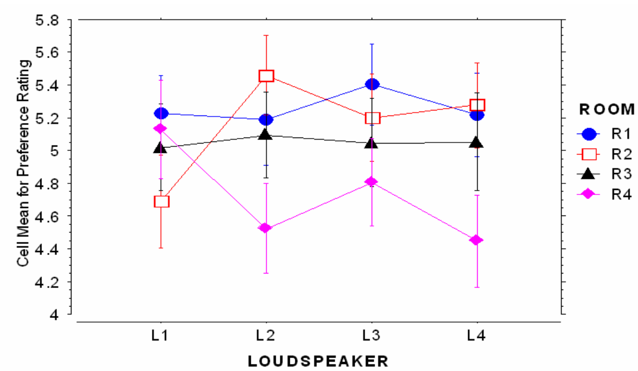


Figure 8 The mean preference ratings and 95% confidence intervals for the interaction between Loudspeaker and Room.

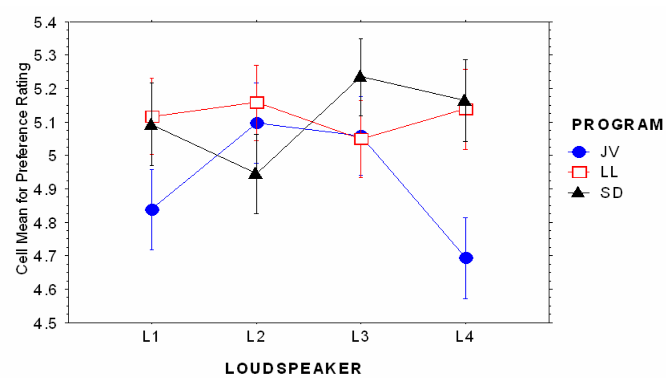


Figure 9 The mean preference rating and 95%

confidence intervals for the interaction Loudspeaker and Program.

3.4. Effects Due to Listening Experience

Previous studies on subjective evaluation of loudspeakers have shown there are significant differences in the performances of trained and untrained listeners [57]-[60]. Together these studies indicate that trained listeners give more reliable and discriminating loudspeaker ratings, yet the overall rank ordering of loudspeaker preferences among trained and untrained listeners are essentially the same [60].

The influence of listener training and experience in this study was examined to determine what role, if any, it had on the reliability and discrimination of preference ratings. A second goal was to determine whether listening experience had any effect on loudspeaker and/or room preferences, and room acoustic adaptation.

All listeners in this study had normal audiometric hearing and passed a spectral identification screening test. However, there were significant differences among them in terms of their years of experience participating in formalized loudspeaker tests. The vast majority (65%) of the listeners were recent recruits having less than 2-6 months experience in formal loudspeaker tests. The remaining subjects (35%) had 2-15 years of experience in formalized loudspeaker listening tests. Using this criterion the subjects were divided into two groups based on whether they had two or more years of experience in loudspeaker tests. Although the two groups were not equal in size, there were an equal number of experienced subjects in both successive and intermixed treatments.

A repeated measures ANOVA of the experimental data was performed with Experience as a between-subjects factor. Experience was significant main effect; $F(1, 21) = 7.4$, $p = 0.01$. Fig. 10 plots the mean preference ratings and upper 95% confidence interval for Experienced (2 or more years of listening test experience) and Inexperienced (< 2 years) listeners.

The graph reveals a familiar scaling effect related to listener experience and training noted in [60]. As found in previous studies, experienced listeners tend to use the lower part of the preference scale and spread their ratings further apart on the scale to register their more critical and discriminating sentiments. The individual listener F-statistic has been used by researchers to assess and compare the discrimination and reliability of listener's loudspeaker ratings [50], [60]. A 2-way ANOVA was performed on each individual listener to calculate their F-statistics for the factors loudspeaker and room. The average individual F-statistics for these factors are shown in Fig. 11 for experienced and inexperienced listeners.

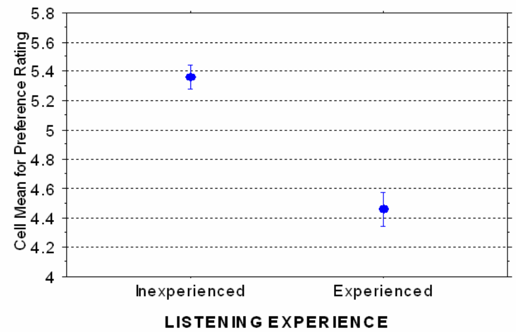


Figure 10 The mean preference ratings and 95% confidence intervals for experienced and inexperienced listeners.

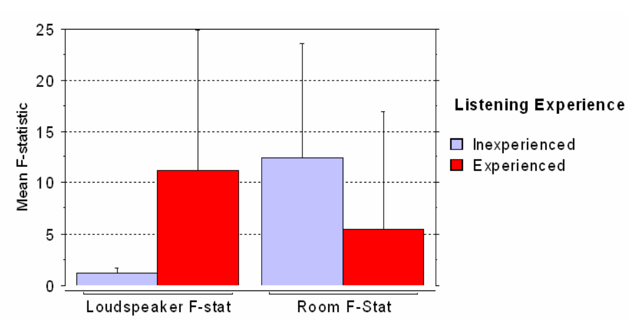


Figure 11 The mean individual listener Loudspeaker and Room F-statistics for Experienced and Inexperienced listeners.

The large confidence intervals, particularly for the experienced subjects reflect the small sample size (i.e. 8 subjects only) and the range of differences in their performances. However, the graph does show an interesting trend. The higher loudspeaker F-statistics for the experienced listeners indicates they were more discerning and/or reliable in rating the loudspeakers compared to the inexperienced listeners.

For the room variable, the opposite trend was found; inexperienced listeners were more discerning and/or reliable when rating room effects than the experienced listeners. This contrast suggests that experienced listeners may have focused more on the loudspeaker differences, while the inexperienced listeners were more focused on room effects.

To explore this idea further, the mean loudspeaker and room preference ratings are plotted for both experienced and inexperienced listeners in Figs. 12 and 13, respectively. As noted earlier, the experienced listeners gave lower preference ratings on an absolute scale. But Fig. 12 also shows they were more discriminating among the loudspeakers than the inexperienced listeners, particularly in regard to L4, which they rated lower than the other three loudspeakers. The inexperienced listeners on the other hand have no significant differences in loudspeaker preferences.

Turning to Fig. 13, the inexperienced listeners were more discriminating among the rooms than the experienced listeners. The inexperienced listeners

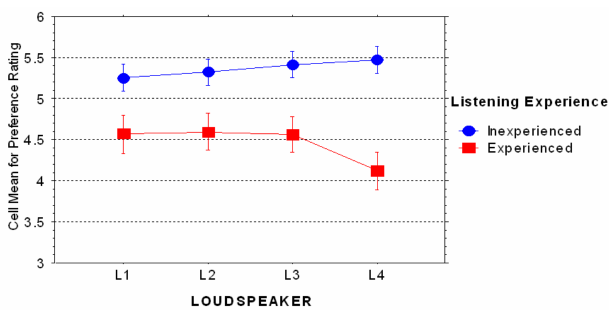


Figure 12 The mean preference ratings and 95% confidence intervals for Loudspeaker plotted for experienced and inexperienced listeners.

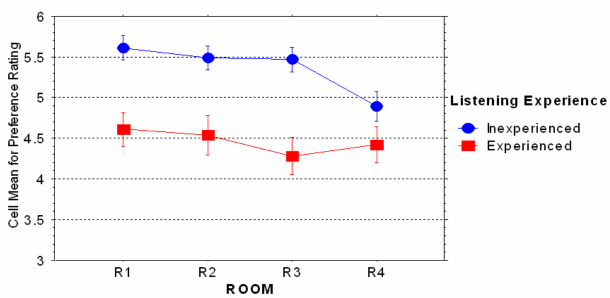


Figure 13 The mean preference ratings and 95% confidence intervals for Room are plotted for experienced and inexperienced listeners.

were much more disapproving of the highly reflective Room R4, than the experienced listeners. The experienced listeners have no strong room preferences, although they tended to give lower ratings in Room R3.

Other studies have reported differences in how trained versus untrained listeners weight the relative importance of timbre and spatial attributes. Ando found trained musicians preferred lower levels of reflected sounds than non-musicians [62]. Rumsey et al., [63] found that in surround sound reproduction listening tests experts tended to weight the timbre and frontal stage spatial fidelity more strongly than untrained listeners, who cared less about these attributes and focused on the envelopment of the surround channels. Rumsey also found similar responses from audio experts who preferred the original stereo material to a 5-channel up-mixed version of it [65]. Evidently untrained listeners are generally more willing to sacrifice some timbral accuracy and frontal stage spatial fidelity for the added sense of envelopment that the lateral energy from the surround speakers or room reflections provides.

The same underlying principles may be at work in this study. If we assume that the respective Loudspeaker and Room effects are timbre versus spatial-related, this could explain why experienced listeners were more influenced by Loudspeaker (i.e. timbre) effects, and inexperienced listeners were influenced by Room (i.e. spatial) effects. The experience gained in years of loudspeaker tests probably helps those listeners focus on the more subtle loudspeaker effects versus the more obvious room effects, which the inexperienced listeners tended to gravitate towards in their preference ratings. Further experimental verification is needed to confirm this observation.

4. DISCUSSION

This section discusses the experimental results in the context of previous related studies. The relationship between the acoustic measurements of the loudspeakers and rooms and their influence on preference is introduced. The influence of room acoustic adaptation on listeners' preferences is also discussed.

Table 8: Contrasts between the current study and Olive et al. [4].

Independent Variable	Olive et al. (1995)	Current Study
Loudspeakers	The three loudspeakers had significant differences in their on-axis (direct sound) and off-axis frequency responses and their low frequency performances.	The four loudspeakers had virtually the same on-axis and low frequency responses The only differences were their off-axis response above 300 Hz.
Loudspeaker Position	Three different positions; positional effects were balanced by comparing each loudspeaker in each position.	Loudspeakers were compared in identical positions in each room.
Mono versus multiple channels	Single channel (center) mono comparisons.	5-channel surround comparisons
Rooms	3 rectangular rooms and an L-shaped room, ranging in volume from 36-91 m^3.	All 4 rooms are rectangular ranging in volume from 60 – 151 m^3. More variation in reflectivity.
Trial Ordering	Successive treatments of room, a between-room trial structure where each loudspeaker is compared among four rooms within same trial.	Successive and Intermixed treatments of rooms
Program	4 selections of monophonic pop and classical music.	3 selections of 5-channel surround sound pop and jazz vocalists.
Playback Method	In situ and binaural recordings reproduced through headphones without binaural room scanning or head-tracking	Binaural room scanned BRIR's reproduced through headphones equipped with a head-tracker

4.1. Comparison of Results to Earlier Studies'

Very few studies have experimentally examined the influence of room acoustic variations on loudspeaker preferences [3]-[6]. These studies used monophonic signals reproduced through single loudspeakers that differed in terms of their directivities and/or frequency responses measured on and off-axis. These are known to be important loudspeaker parameters that were not adequately controlled to allow the researchers to identify what acoustic factors listeners were responding to. Most of theses listening tests were done in situ, rather than using binaural playback method; this raises the issue of whether the effect of the room and its interaction with loudspeaker and other variables were biased by the listeners' knowledge of which room they were in during the test. In contrast to the previous studies, our study carefully controlled important loudspeaker parameters and loudspeaker positional effects within the room. The BRS method ensured all independent test variables such as the room, loudspeaker and its position in the room were evaluated in a truly controlled, double-blind manner.

The authors are only aware of one previous study [4] that is closely related to the current one. There are some important differences between them, in terms of their experimental design, method and the acoustical features of the stimuli that are summarized in Table 8. In the previous study, listeners gave preference ratings for three different loudspeakers in four different rooms that were evaluated both in situ and using binaural recordings reproduced through headphones. The different rooms were blocked as successive treatments. The experimental results showed the loudspeakers were the dominant factor in listeners' preference ratings for both in situ and binaural results where successive room treatments were evaluated. The room variable had no effect on listeners' ratings, which the authors in [4], and later Toole [1], attributed to room acoustic adaptation.

In contrast to those findings, the current study found different room acoustics had an effect on listeners' preference ratings, and Loudspeaker was not a main effect. The lack of significant Loudspeaker effects observed in the current study is likely due to three reasons:

1. The acoustical differences among the loudspeakers were sufficiently small that the majority of listeners could not reliably formulate loudspeaker preferences.

2. The loudspeaker comparisons were made in multichannel, not mono or stereo, as in previous studies [3]-[6], which may have diminished listeners' ability to reliably discriminate among the loudspeakers.

3. The strong interactions between the loudspeakers and rooms tended to obfuscate any main effects due to loudspeaker (see section 4.2).

The acoustic differences among the loudspeakers were carefully controlled so that listeners heard a constant direct sound from all loudspeakers, in all rooms. The only acoustic differences among the loudspeakers were related to their off-axis responses, and these colorations were only contained within the reflected sounds that arrived at the listener. In contrast to this, the loudspeakers in the previous study [4] had significant differences in their amplitude responses measured on-axis, off-axis and at low frequencies (see fig. 2 in [4]). All three are important loudspeaker parameters used to predict loudspeaker preference ratings in reflective rooms [6], [42], [43], [54], [55].

The second possible explanation for the lack of strong loudspeaker effects in this current study is more speculative since the hypothesis has not yet been tested. Toole has reported that listeners tend to be less discriminating when comparing loudspeakers in stereo compared to mono [55], [65]. While the overall rank order of the loudspeakers tends to be the same, listeners are more forgiving of loudspeakers auditioned in stereo, that when auditioned in mono received lower ratings. According to Toole, loudspeaker discrimination in stereo vastly improves if the stereo recordings have strong monophonic components that are commonly found in multi-track recordings made with single microphones. More recently, Toole has attributed the mono-versus-stereo loudspeaker effect to listeners making trade-offs between the loudspeaker's timbral accuracy and its perceived spaciousness [66]. Listeners evidently overlook the loudspeaker's off-axis colorations if the loudspeaker has wider dispersion (i.e. lower directivity). The wider dispersion loudspeakers produce stronger lateral reflections in the room that generate a greater degree of listener envelopment or spaciousness. Any loudspeaker advantage in terms of its spaciousness becomes become nullified when the loudspeakers are auditioned using stereo recordings with a strong out-of-phase difference component. At this point, the loudspeaker's timbral accuracy may become the determining factor in listeners' loudspeaker preferences. How this effect operates with multichannel loudspeakers and recordings is something that requires experimental verification.

4.2. Influence of Loudspeaker on Preference for Reproduced Music

The influence of loudspeaker on preference ratings was only exhibited via its interaction with the room. That is to say, there was no main effect, but only a change in the pattern of loudspeaker preferences due to the loudspeaker interactions with the room in which they were positioned.

The interaction between loudspeaker and room acoustics is shown again in Fig. 14. There are three main points to be made:

1. Certain listening rooms (R4 and R2) were better than others in revealing the audible presence of off-axis colorations in the loudspeakers based on the larger range of loudspeaker preference ratings observed in those particular rooms.

2. Highly-reflective rooms, like Room R4 produced lower preference ratings.

3. The loudspeakers with the best off-axis frequency responses (L1 and L2) were sometimes rated very high or very low in rooms (R4 and R2) good at revealing the audible presence of loudspeaker off-axis colorations.

The poorest listening rooms for revealing off-axis performance differences in loudspeakers were the R3 and R1. In these two rooms, there were no significant differences in the loudspeaker preference ratings. Evidently there was an acoustic feature common to these two rooms that made them particularly poor at revealing off-axis spectral colorations. Room R3 had a higher proportion of direct-to-reflected sound at the listener location that made off-axis loudspeaker colorations more difficult to discern. In room R3, the side walls were further away from the loudspeakers and listeners, so that the important lateral reflections arrived later and were lower in level than the reflections in the other three rooms. Room R1, the smallest room in this study, required the listener to be only 1.83 m from the five loudspeakers, and this may have produced a higher proportion of direct-to-reflected sound than found in Room R2.

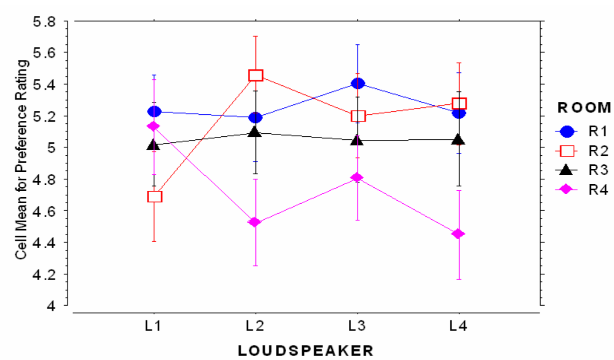


Figure 14 The mean preference ratings and 95% confidence intervals for the Loudspeaker-Room interaction.

Rooms that were good for discerning off-axis loudspeaker colorations were Rooms R1 and R2. Room R4 was the most reflective room in the study, and Room R2, while not as reflective, still had higher levels of early lateral reflections than found in Rooms R3 or Room R1.

The previous loudspeaker-room study reported similar loudspeaker-room interactions (see section 3.5 of [4]), where certain rooms were better than others at revealing audible differences among loudspeakers. Room A, the smallest and most acoustically dead room in that study, produced the smallest differences in loudspeaker preferences. The authors attributed this to the lack of room reflections at mid-high frequencies where the loudspeaker colorations were present. Rooms L was among the best room at revealing loudspeaker preferences:

"In the case of Room L, speaker-position interactions were not a factor. Therefore, there was likely some acoustical characteristic about the room that made audible differences between the speakers subjectively more apparent. Perhaps, it was a combination of the room's liveness and asymmetry that produced a greater variety of reflections arriving at different angles of incidence and times. This would have certainly placed more extreme demands on the off-axis performances of the loudspeakers, which the measurements indicate were as different as their on-axis ones." (p. 13)

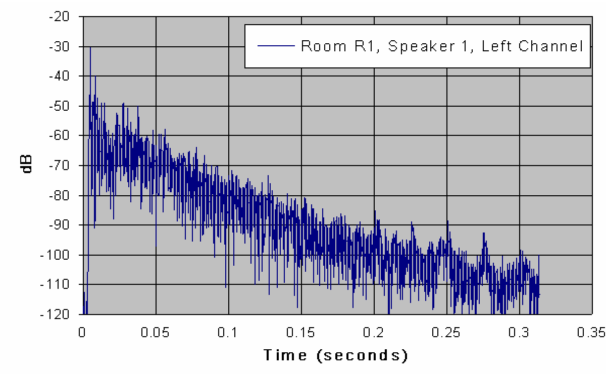
The second observation regarding Fig. 14 is that the preference ratings in Room R4 were generally lower than in any of the other rooms. This could be related to two factors: 1) the audible colorations in the loudspeakers produced lower preference ratings for multichannel music reproduction and/or 2) the higher level reflections in Room R4 diminished the clarity, and possibly other important perceptual attributes of the reproduced music. This finding underlines the importance of testing loudspeakers in rooms that are sufficiently reflective so that the audibility of off-axis problems in the loudspeakers can be assessed. If listeners prefer the loudspeakers in such demanding listening rooms, there is a greater likelihood that the loudspeakers will be accepted in a variety of different, real-world listening rooms.

The third observation regarding Fig. 14 is related to correlations between the objective and subjective performances of the loudspeakers among the different rooms. One would expect Loudspeakers L1 and L2 to receive similar preference ratings given the similarity in their off-axis frequency responses. However, in the rooms that are audibly revealing of loudspeaker off-axis colorations (Rooms R2 and R4) L1 and L2 tended to be rated quite wide apart on the preference scale. Clearly the room and loudspeaker are interacting in a way that cannot be explained alone by the anechoic loudspeaker measurements. Analysis of the in-room measurements are currently underway, and will hopefully unravel this mystery.

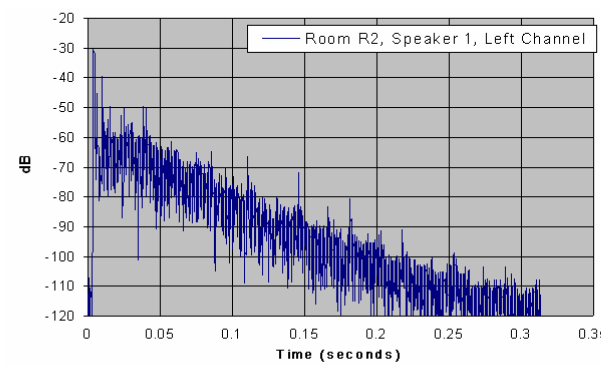
4.3. Room Reverberation is a Factor in Listeners' Preference for Multichannel Music Imagery

There was a significant main effect due to Room. A post-hoc test revealed that the other three listening rooms were preferred over Room R4. There were no statistically significant preferences among the other three rooms. This section examines how Room R4 was acoustically different from the other rooms to possibly explain its lower preference ratings.

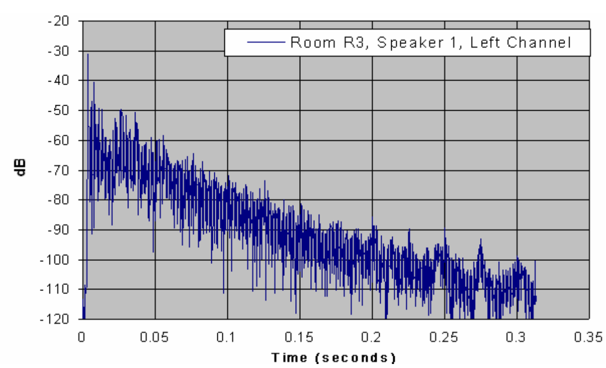
The acoustic differences among the four listening rooms in terms of their reflective energy are plotted in Figs. 15 (a)-(d). Shown are the energy-time curves (ETC) of each room. Normally, the ETC is calculated from the impulse response [66], [67], but here the BRIR (left ear only) was used. This means that the ETC's shown here contain the frequency response and directional properties of the binaural manikin. The BRIR was measured at the listening location using loudspeaker L1 located in the left front position. Only the first 300 msec of the ETC are shown, since most of the energy at that point, has decayed. The energy decay patterns are very similar among Rooms R1, R2 and R3, although there are some small differences in the fine details of the ETC. The similarities in their ETC's may partially explain why the loudspeaker preference ratings did not vary significantly among these three rooms. On the other hand, Room R4 has significantly higher levels of late



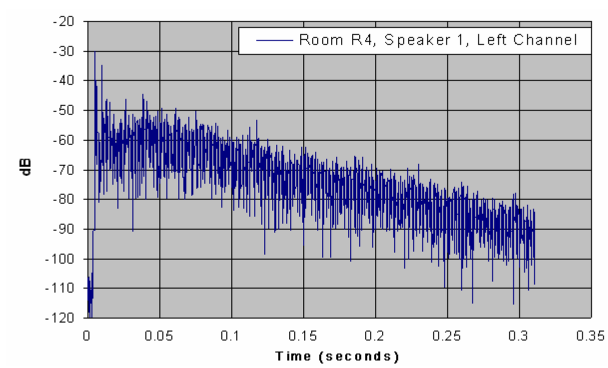
(a)



(b)



(c)



(d)

Figure 15 The energy-time curves (ETC) for Rooms: (a) R1, (b) R2, (c) R3, and (d) R4. The ETC is calculated from the BRIR (left ear only) measured at the listening location using the left front loudspeaker.

reflected energy beginning at 60 msec that are 15-20 dB higher than those found in the other three rooms. This later arrival, higher level reflections likely contributed to its lower preference ratings. While late arrival lateral reflections (>80 msec) produce higher values of listener envelopment (LEV) [69] - a desirable attribute for concert halls [64], [65] - they can also produce as loss in clarity due to increased forward temporal masking [[71]. To verify the exact physical cause of the lower ratings in Room R4, additional experiments would be required that systematically manipulated the reflective properties within the room. In addition to the ETC, we are looking into other acoustical measurements of the room and loudspeaker that might further explain the influence of the loudspeaker-room interactions on listener preference.

A formal scientific investigation into the underlying perceptual attributes of the different listening rooms has not yet been done. Informal comments solicited from listeners revealed that Room R4 sounded "excessively reverberant, very reflective, overly live, and too bright, with emphasis on the midrange frequencies". These comments are related to the clarity and spectral balance of the auditory imagery, and may be linked to its lower ratings. It is interesting to note the general absence of criticisms related to early reflections such as image shift and image widening (ASW), confirming previous studies [1], [72] that show these reflections are below the level where these effects occur.

4.4. Comparison of Room Effects with Previous Studies

Bech evaluated four loudspeakers in three different rooms, and found that loudspeaker ratings tended to be higher in one of the rooms [3]. There were no in-room measurements of the loudspeakers offered to possibly explain the nature of the effect, but reverberation measurements of the rooms were provided. The preferred room had a relative consistent RT_{60} of 0.4 s between 125 Hz and 8 kHz, which is the average reverberation for a domestic room [38]. The second room had significantly less reverberation (RT_{60} = 0.25 s), whereas the third room had significantly more reverberation between 160 Hz-500 Hz.

The previous loudspeaker/room study [4] discussed earlier in this paper, reported no significant room effects. ETC's of the rooms were not provided so direct comparisons betweens studies are not possible. However, the authors described the rooms' physical dimensions, acoustic treatment and their reverberation characteristics. All four rooms fall within the reverberation times found in domestic homes: RT_{60} of 0.4 s with a standard deviation of 0.1 s from 80 Hz to 4 kHz [38]. The authors concluded:

"It is sufficient to say that all of the four rooms meet or closely approach some, or all of the IEC specifications, and that none are atypical of average real-world domestic rooms" (p. 4).

There were significant differences in the volumes of rooms between the two studies summarized in Fig. 16. The rooms in the current study were much larger due to the larger space requirements for multichannel loudspeaker setups.

Room volume alone does not seem to be a good indicator of the influence of room on loudspeaker preferences, since discrepancies in room volumes can be compensated for with additional amounts of

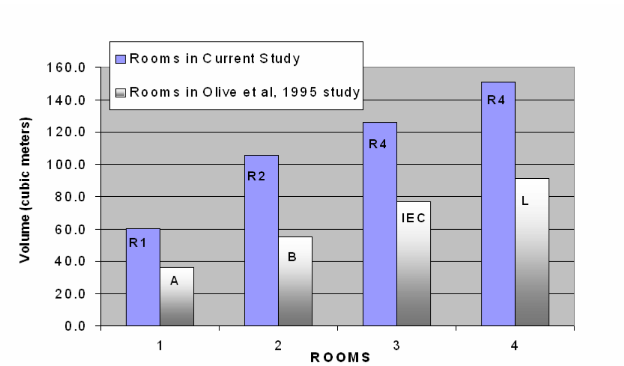


Figure 16 The volumes (m^3) of the four listening rooms in used in the current study and the previous study [4].

absorption. A more important factor is the room's reflectivity, particularly the level, arrival time and spatial distribution of the reflections. As Toole has pointed out in [1], most domestic rooms do not have diffuse sound fields due to the directivity of the loudspeakers, the low ceiling, and the proportionally large amount of absorption and scattering objects in the room and along its boundaries. As a result, the late reflected sound field is great diminished. However, in the current study the additional loudspeakers distributed about the room, would have increased amount of room reflections compared to those produced from a single loudspeaker.

4.5. The Influence of Program on Loudspeaker and Room Preferences

No significant program effects were reported in the current or previous related studies. This was not unexpected since the programs were carefully selected on the basis of their homogeneous spectral and spatial characteristics.

In the current study, there was a small but statistically significant interaction between program and loudspeaker. The interaction was largely isolated to the female jazz vocalist Program (JV) and Loudspeaker L4, which had the poorest off-axis response. Most likely the loudspeaker colorations were more audible for this program. To confirm this supposition, a separate experiment would be required to measure the intensities of the perceptual attributes of each loudspeaker for each program.

Program interactions can also occur when the program itself contains a distortion that is complementary to a similar, but opposite one in the loudspeaker. For example, a “bright” program can make a “dark” loudspeakers sound neutral.

Interactions between the directivity of the loudspeaker and the “genre” of music have also been reported. Klippel found listeners preferred more directional loudspeakers for speech, and less directional speakers for music [6]. Similar program interactions with room acoustics have been found as well. In a study on the preferred acoustics of professional control rooms Voelker [75] found more reflective control rooms (RT_{60} = 0.7s) were preferred for chamber and organ music, and less reflective rooms (RT_{60} = 0.4 s) were preferred for pop and disco music. A nearly reflection free room (RT_{60} = 0.2 s) was preferred for solo drum music. The notion that certain acoustic spaces are preferred for different genres of music is well-known among concert hall designers, and music composers have even written music for specific performance spaces in mind.

A more recent study by Weisser and Rindel [75] also found that the preferred reverberation characteristic in small rooms depends on the program. Listeners made comparative ratings of a loudspeaker reproducing monophonic programs in 7 different small rooms via binaural recordings reproduced over headphones. Subjective ratings included overall sound quality, boominess and boxiness (i.e. coloration) or coloration. The preferred reverberation time for speech were a low as possible, and 0.3 – 0.5 s for music. Objective metrics based on the rooms’ 1/3-octave reverberation times could predict the subjective ratings with reasonably good accuracy.

These types of program interactions between loudspeakers and rooms are hopefully limited to monophonic and stereophonic reproduction. With the advent of multichannel audio and improvements in recording methods, interactions between programs, loudspeakers and rooms will hopefully disappear in the future. The spatial distribution of more loudspeakers should provide better control and delivery of the direct-to-reflected sounds, which can be easily tailored to suit the music genre during the production of the recording. Multichannel audio will eliminate the need to build special listening rooms for specific genres of music. In this current study we found no significant program interactions with room, a hint perhaps that multichannel setups are more conducive to a wider range of listening rooms and musical styles.

4.6. Did Room Acoustic Adaptation Occur in this Study?

In this study, the differential exposure to the different room acoustics failed to produce a significant effect on listeners’ preference ratings. Therefore, there is no direct statistical evidence to prove that room acoustic adaptation occurred. The lack of statistical evidence doesn’t necessarily mean that room acoustic adaptation did **not** occur. Most likely, the time course over which the intermixed treatments were evaluated was too long for listeners to avoid adaptation to the room acoustics.

In the previous study [4], evidence for room acoustic adaptation was based on a different paradigm than what was used in the current study. In the previous study, the break-down in room acoustic adaptation occurred when listeners made instantaneous direct comparisons among the four rooms via a binaural playback device, which produced significant room effects that had not occurred in the successive treatments of the listening rooms. The direct multiple comparisons of rooms meant listeners literally had just a few seconds to adapt to the room acoustics before the room was changed. Room differences would have been perceptually highlighted since the room was the changing feature within the trial, while the loudspeaker remained constant. This time course was much shorter than the 1-2 minutes time course used for evaluating the intermixed treatments in this study. This gives further credence to the notion that room acoustic exposure times must be significantly reduced in order to produce a break-down in room acoustic adaptation.

There is additional evidence that suggests room acoustic adaptation may have been working in the current study. The influence of the method and room is plotted in Fig. 17. Even though the interaction is not statistically significant [F (3, 1) = 1.19, p = 0.32], it is shown to see whether further investigation is warranted.

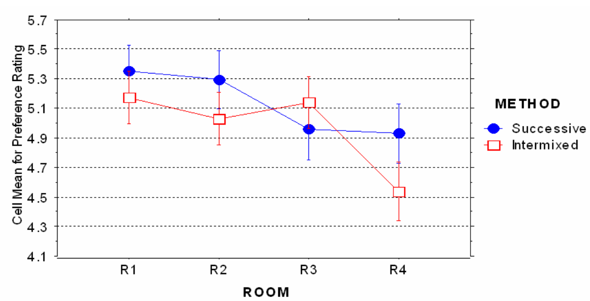


Figure 17 The mean preference ratings and 95% confidence intervals for the four listening rooms for the successive and intermixed treatments.

There are two notable trends. First, the mean preference ratings related to Room tended to be lower for the intermixed treatments. This pattern was true for three of the four rooms. Secondly, for Room RR, the difference in mean ratings for the successive and intermixed treatments is almost statistically significant. The comparatively lower ratings given in the intermixed treatments suggest that listeners may have adapted less to the rooms compared to successive treatment group of listeners. This effect was most pronounced in Room R4. More investigation is needed to confirm whether this trend is real or not.

The efficiency and speed in which listeners adapt to different acoustic environments may seem surprising at first glance, until one considers that we adapt to room acoustics every day, often without being aware of it. As we walk through different acoustical spaces at home or at work on daily basis, their acoustical properties can vary significantly. Yet we hardly notice the effect of these rooms on the timbre, spatial properties and intelligibility of the people we are conversing with. The timbral and localization constancy of sound sources within these different reflective spaces indicates that powerful, room-compensating perceptual processes are at work. It is only when we move into in a large reflective space that we become aware of how the room changes the timbre, intelligibility and spatial aspects of sounds around us.

5. CONCLUSIONS

This paper describes an experiment that measured how loudspeaker preferences for multichannel music imagery are influenced by different listening room acoustics, and the extent to which listeners adapt to them. Four surround loudspeaker systems were evaluated in four different listening rooms using three different 5-channel music selections. The four loudspeakers differed only in their off-axis frequency response above 300 Hz, meaning listeners heard loudspeaker variations based only on their reflected sounds. Two groups of listeners were given differential exposure to the listening room acoustics via a BRS measurement and playback system. One group of listeners gave loudspeaker preference ratings where the room was held constant throughout the listening session, and only changed in subsequent sessions (successive treatments). A second group of listeners did the same task except the listening room was changed every subsequent trial within the listening session (intermixed trials).

The results were analyzed as a 2 x 4 x 4 x 3 x 3 repeated measures ANOVA where the independent variables were Trial Order Method (2 levels), Loudspeaker (4), Room (4), Program (3) and Observation (3), and the dependent variable was preference rating. Trial Order Method was treated as a between-subjects factor. All statistical tests were performed at an alpha level of 0.05. The significant effects and interactions are summarized below.

Room reflectivity influences loudspeaker preferences - Room acoustics had a significant effect on listener loudspeaker preference ratings. The lowest preference ratings were observed in the highly-reflective room (Room R4), which was less preferred over the other three listening rooms. There were no significant differences in preference ratings among the other three listening rooms.

Influence of loudspeaker on preference for multichannel reproduction of music – The influence of loudspeaker on preference ratings was only exhibited via its interaction with the room. That is to say, there was no main effect, but only a change in the pattern of loudspeaker preferences due to the loudspeaker interactions with the room in which they were positioned. The reflectivity of the room is an important determining factor in loudspeaker preferences. There were significant loudspeaker preferences in the more reflective Rooms (R4 and R2), and none in the Rooms (R1 and R3) where listeners were in a more direct sound field. The audibility of the different off-axis frequency response irregularities in the loudspeakers were clearly enhanced by rooms that provided higher levels of early reflections. Loudspeakers with more irregular off-axis frequency responses tended to score lower in the more reflective rooms.

Program influences loudspeaker preference – A significant interaction was found between Program and Loudspeaker, which was largely isolated to Program JV (Jazz band with female vocal). This program produced lower ratings for loudspeakers L1 and L4. The exact cause of this interaction is being investigated

A listener's experience influences how they scale their preferences and weight the relative importance of loudspeaker and room effects – Prior experience in loudspeaker tests was a factor in how listeners scaled their preference ratings, and the relative influence the factors Loudspeaker and Room each had on their preferences. Listeners with +2 years of experience had, on average, higher Loudspeaker F-statistics than the inexperienced listeners indicating that experienced listeners were more discerning of loudspeaker differences. On the other hand, the inexperienced listeners had higher Room F-statistics indicating they were more discerning of room effects than the experienced listeners. We suspect this contrast between experienced and inexperienced listeners is related to differences in how they weight the relative importance of timbre (loudspeaker) versus spatial (room) effects in multichannel audio reproduction. More experimental verification is needed to confirm whether this is true. Experienced listeners, on average, used lower preference ratings as has been reported in other studies [60].

Room preference for multichannel music reproduction and its relation to acoustic measurements

(a) *Above the transition region* – Room preferences related to acoustic effects above the transition frequency (300 Hz) can be partially explained by the energy-time curve (ETC) of the BRIR (left ear only). The room that received the lowest preference ratings had significantly more energy (15-20 dB) in the ETC between 60-300 msec than the other three rooms.

(b) *Below the transition region* – Loudspeaker-room interactions below the room transition region (about 300 Hz) were not reported in this paper. Analysis of these effects will be the topic of future paper.

Room acoustic adaptation – The differential exposure to the different room acoustics failed to produce a significant effect on listeners' preference ratings. Therefore, there is no direct statistical evidence to prove that room acoustic adaptation occurred or that it didn't occur. Most likely, the time course over which the intermixed trials were completed was too long to prevent room acoustic adaptation from occurring. If this is true, listeners appear to adapt very quickly to the listening room acoustics. Comparing room preference ratings for successive and intermixed treatments provides evidence that there was less adaptation in the intermixed treatments because the preference ratings were generally lower, and there were larger room effects than observed with the successive treatments.

5.1. Generalizability of this Study

The experimental findings of this study may or may not apply to listening conditions outside those tested here. The generalizability of a study, often referred to as its external validity, describes the degree to which the findings will be true over a range of conditions outside those tested in the laboratory. Below are some possible factors in this study that may limit its generalizability or external validity.

Rooms – The listening rooms in this study were all simple rectangular-shaped rooms, with dimensions that fall within a range real-world values. While rectangular rooms are very common, some domestic listening rooms have irregular shapes, doors and windows, open floor-plans, and cathedral ceilings higher than those tested here. These types of rooms need to be included in future studies.

Acoustical Treatment of Room – Three of the rooms in this study had reverberation and acoustical treatments that fall within average domestic rooms [38]. Room R4 was unusually reverberant, and is likely outside the range of most domestic rooms of its size.

Loudspeakers - A conscious decision was made not to include poor quality loudspeakers to avoid possible stimulus recognition biases that might swamp any room effects and room acoustic adaptation. All four loudspeakers in this study were probably among the best examples of their kind sold today. Speakers with unusual directivities were not represented in this study.

Loudspeaker Setups – A limiting factor in the generalizability of this study was the use only ITU-R BS.775-1 [33] setups. Perfect, symmetrically arranged circles of equidistant loudspeakers are not so common in consumer's homes, since practical, aesthetic or spousal-related considerations usually take priority. ITU-R setups are most commonly used by audio academics, scientists and serious audiophiles. More data is needed for loudspeaker setups that are arranged asymmetrically, diagonally in a corner, with different angular (vertical and horizontal) positioning of the surround speakers.

Programs – The temporal and spectral envelope of the program affects the detection of reflections [1], [70] and resonances [71], and other spectral irregularities in loudspeakers and rooms [1]. This study used three selections of pop and jazz vocal music that contained both temporally continuous and discontinuous signals. Not included in this study were highly impulsive signals such as clicks or percussive transients that produce lower detection thresholds for late arrival reflections [1], [70]. These signals produce much lower thresholds for late arrival reflections. In highly reflective rooms, adaptation may diminish for discontinuous, impulsive-type signals.

Room Acoustic Adaptation – Factors that may produce different amounts of room acoustic adaptation than were measured in this study include: the nature and complexity of the listening task, the physical properties of test signals, the number of loudspeakers used, the listeners' training and experience, the exposure time to the room, and the acoustic properties of the loudspeakers and rooms. Shorter time courses (< 30 s) than used in this study are recommended to produce less room acoustic adaptation.

6. ACKNOWLEDGEMENTS

This work was supported by Harman International. The authors would like to thank members of the Harman International R&D Group for their contributions to this paper; Todd Welti for his work on the BRS calibration and measurement system; Allan Devantier who engineered the adjustable loudspeaker; Eric Hu who customized the BRS playback software; John Jackson who along with Todd Welti helped gather the +6,000 sets of BRIR and run over 100 listening test sessions; David Hunt and Sean Hess who participated as listeners and assisted in the development of listening test software; Alex Miller who helped measure the loudspeakers; Floyd Toole whose work in loudspeakers and rooms has stimulated much of the motivation behind this work. The authors also thank Tim Nind, Phil Muzio and Mae Nutley at Harman Becker Automotive Systems for their support and discussions related to this work. Kevin Voecks and Floyd Toole also provided useful comments on the text. Finally, we thank the numerous listeners who participated in these tests.

7. REFERENCES

[1] F.E Toole, "Loudspeakers and Rooms for Sound Reproduction – A scientific Review," J. Audio Eng. Soc., vol. 54, no. 6, pp. 451-476 (July/August 2006).

[2] S. Olive, P. Schuck, S. Sally and M. Bonneville, "The Effects of Loudspeaker Placement on Listener Preference Ratings," J. Audio Eng. Soc., vol. 42, no. 9, pp. 651-669 (September 1994).

[3] S. Bech, "The Influence of the Room and of Loudspeaker Position on the Timbre of Reproduced Sound in Domestic Rooms," Audio

Eng. Soc. 12th International Conference: The Perception of Reproduced Sound, paper 12-007. S. (1993).

[4] S. Olive, P. Schuck, S. Sally, M. Bonneville, K. Momtahan and E. Verreault,"The variability of loudspeaker sound quality among four domestic sized rooms", in the proceedings of the 99th Convention of the AES, preprint 4092 (October 1995).

[5] P. Schuck, S. E. Olive, J. Ryan, F. Toole, S., Sally, S. Bonneville, M., Verreault, E., Momtahan, K., "Perception of Reproduced Sound in Rooms: Some Results from the Athena Project", pp.49-73, in the proceedings of the 12th International Conference, Audio Eng. Soc. (June 1993).

[6] W. Klippel, "Multidimensional Relationship between Subjective Listening Impression and Objective Loudspeaker Parameters," Acustica, vol. 70, pp. 45-54 (1990).

[7] S. Bech, "Timbral Aspect of Reproduced Sound in Small Rooms, II," J. Acoust. Soc. Am., vol. 97 (March 1995).

[8] S. Bech, "Spatial Aspect of Reproduced Sound in Small Rooms, II," J. Acoust. Soc. Am., vol. 103 (June 1996).

[9] H. Wallach, E.B. Newman and M.R. Rosenweig, "The precedence effect in sound localization," J. Exp. Psychol. vol. 27, pp. 339-368 (1949).

[10] H. Haas, "The Influence of a Single Echo on the Audibility of Speech," Ph.D. dissertation, University of Gottingen, Gottingen, Germany (1949); translation reprinted in J. Audio Eng. Soc., vol. 20, pp. 146–159 (March 1972).

[11] J. P. A Lochner and J. F. Burger," The Subjective Masking of Short Time Delayed Echoes by Their Primary Sounds and Their Contribution to the Intelligibility of Speech," Acustica, vol. 8, pp. 1–10 (1958).

[12] P. M. Zurek, "Measurements of Binaural Echo Suppression," J. Acoust. Soc. Am., vol. 66, pp. 1750–1757 (1979).

[13] K. Saber and D.R. Perrott,"Lateralization thresholds obtained under conditions in which the precedence effect is assumed to operate," J. Acoust. Soc. Of Am. vol. 87, pp. 1732-1739 (1990).

[14] B. Rakerd and W.H. Hartman, "Precedence Effect with and without interaural differences - Sound localization in three planes," J. Acoust. Soc. Am., vol. 92, pp. 2269 (1992).

[15] R.K. Clifton, R.L. Freyman, R.Y. Litovsky and D. McCall," Listeners' expectations about echoes can raise or lower echo threshold," J. Acoust. Soc., Am., vol. 95, pp. 1525-1533. (1994).

[16] B. Rakerd and W.H. Hartman, "More on the precedence effect in three planes," J. Acoust. Soc Am.., vol. 95, pp. 2917 (1994).

[17] J. Blauert, *Spatial Hearing - Revised Edition: The Psychophysics of Human Sound Localization*, MIT Press, 1996.

[18] R.K. Clifton and R.L. Freyman, "The precedence effect: Beyond echo suppression," in *Binaural and Spatial Hearing*, edited by R. Gilkey and T. Andersen (Erlbaum, Hillsdale, New Jersey). (1996).

[19] W.M. Hartmann, "Listening in a room and the precedence effect," in *Binaural and Spatial Hearing,* edited by R. Gilkey and T. Anderson, Erlbaum, Hillsdale, New Jersey (1996).

[20] T. Djelani and J. Blauert, "Investigations into the Build-up and Breakdown of the Precedence Effect," Acta Acustica— Acustica, vol. 87, pp. 253–261 (2001).

[21] R. Y. Litovsky and B. G. Shinn-Cunningham, "Investigation of the Relationship among Three Common Measures of Precedence: Fusion, Localization Dominance, and Discrimination Suppression," J. Acoust. Soc.Am., vol. 109, pp. 346–358 (2001).

[22] J. Blauert, J. Mourjopoulos and J. Buchholz, "Room Masking: Understanding and Modelling the Masking of Reflections in Rooms," Preprint

Number: 5312 presented at the 110th Audio Eng. Soc.Convention (April 2001).

[23] J. Blauert and P. L. Divenyi, “Spectral Selectivity in Binaural Contralateral Inhibition,” Acustica, vol. 66, pp.267–274 (1988).

[24] B. G. Shinn-Cunningham, “Localizing Sound in Rooms,” in Proc. ACM SIGRAPH and EUROGRAPHICS Campfire: Acoustic Rendering for Virtual Environments (Snowbird, Utah,2001),

[25] B.G. Shinn-Cunningham, “Learning reverberation: Considerations for spatial auditory displays,” in Proceedings of the International Conference on Auditory Display, Atlanta, GA., pp. 126-134 (April 2-5, 2000).

[26] B.G. Shinn-Cunningham, N.I. Durlach, and RM Held, “Adapting to supernormal auditory localization cues I: Bias and resolution,” J Acoust. Soc Am, vol. 103, pp. 3656-3666. (1998).

[27] B.G. Shinn-Cunningham, N.I. Durlach, and R.M Held, “Adapting to supernormal auditory localization cues II: Constraints on adaptation of mean response,” J Acoust Soc, Am, vol. 103, pp. 3667-3676 (1998).

[28] M. Schoolmaster, N. Kopo, and B. G. Shinn-Cunningham, “Effects of Reverberation and Experience on Distance Perception in Simulated Environments” (Abstract), *J. Acoust. Soc. Am.*, vol. 113, p. 2285 (2003).

[29] M. Schoolmaster, N. Kopo, and B. G. Shinn-Cunningham, “Auditory Distance Perception in Fixed and Varying Simulated Acoustic Environments” (Abstract), J. Acoust. Soc. Am*.*, vol. 115, p. 2459 (2004).

[30] J. Blauert and P. L. Divenyi, “Spectral Selectivity in Binaural Contralateral Inhibition,” Acustica, vol. 66, pp.267–274 (1988).

[31] A. J. Watkins, “Central, Auditory Mechanism of Perceptual Compensation for Spectral-Envelope Distortion,” J. Acoust. Soc. Am*.*, vol. 90, pp. 2942–2955 (1991).

[32] A. J. Watkins and S. J. Makin, “Some Effects of Filtered Contexts on the Perception of Vowels and Fricatives,” J. Acoust. Soc. Am., vol. 99, pp. 588–594 (1996).

[33] ITU-R BS.775-1, “Multichannel Stereophonic Sound System With and Without Accompanying Picture,” International Telecommunications Union, Geneva, Switzerland (1994).

[34] The Recording Academy's Producers and Engineers Wing, “Recommendations for Surround Sound Production", published by the National Academy of Arts and Sciences, 2004. see http://www.grammy.com/Recording_Academy/Producers_And_Engineers/Guidelines/

[35] U. Horbach, A. Karamustafacglu, R. Pellegrini, P. Mackensen and G. Thiele, G.," Design and Applications of a Data-based Auralization System for Surround Sound,” 106th Convention, Audio Eng. Soc., preprint 4976 (April 1999).

[36] S. Olive, T. Welti, W. Martens, ”Listener Loudspeaker Preference Ratings Obtained In Situ Match those Obtained Via a Binaural Room Scanning Measurement and Playback System,” presented at the 122nd Audio Eng. Soc., Convention, 122nd Vienna, preprint 7034 (May 2007).

[37] S. Bech, W. Ellermeier, J. Ghani, M-A. Gulbol, and G. Martin,” A Listening Test System for Automotive Audio - Part 2: Initial Verification," Preprint Number: 6359 Convention: 118 (May 2005).

[38] Bradley, J.S., "Acoustical Measurements in Some Canadian Homes", vol. 14, pp. 19-25 (Oct. 1986)

[39] RPG, see http://rpginc.com/products/skyline (2007)

[40] S. E. Olive, B. Castro, and F. E. Toole, “A New Laboratory for Evaluating Multichannel Audio Components and Systems,” presented at the 105th Convention of the Audio Engineering Society, *J. Audio Eng. Soc.* (*Abstracts*), vol. 46, pp. 1032, 1033 (1998 Nov.), preprint 4842.

[41] Francis Rumsey "Spatial Audio and Sensory Evaluation Techniques – Context, History and Aims," in Proc. of Spatial Audio & Sensory Evaluation Techniques, Guildford, UK (April 2006).

[42] S. Olive, "A Multiple Regression Model for Predicting Loudspeaker Preference Using Objective Measurements: Part I—Listening Test Results," presented at the116th Convention of the Audio Engineering Society, J.Audio Eng. Soc. (Abstracts), vol. 52, p. 861 (2004 July/Aug.), convention paper 6113.

[43] S. Olive, "A Multiple Regression Model for Predicting Loudspeaker Preference Using Objective Measurements: Part II—Development of the Model," presented at the 117th Convention of the Audio Engineering Society, J.Audio Eng. Soc. (Abstracts), vol. 53, p. 83 (2005 Jan./Feb.), convention paper 6190

[44] William L. Martens, Marui, Atsushi, and Kim, Sungyoung, "Investigating Contextual Dependency in a Pairwise Preference Choice Task," presented at the AES 28th International Conference, Pitea, Sweden (June 30-July 2, 2006).

[45] S. Kim, M. DeFrancisco, K. Walker, A. Marui and W. Martens, "Listener Preferences in Multichannel Audio: Examining the influence of musical selection on surround microphone technique," in proceedings of AES 28th International Conference, Pitea, Sweden, (July 2006).

[46] S. Zielinski., "On some biases encountered in modern listening tests," in Proc. of Spatial Audio & Sensory Evaluation Techniques, Guildford, UK, (April 2006).

[47] W. L. Martens, "Contextual Effects in Sensory Evaluation of Spatial Audio: Integral Factor or Nuisance?" in proceedings of Spatial Audio & Sensory Evaluation Techniques, Guildford, UK (April 2006).

[48] Geoffrey Keppel and Thomas D. Wickens. Design and Analysis: A Researcher's Handbook. Pearson Education, Upper Saddle River, New Jersey, 4th edition (2004).

[49] 11. N. Kousgaard, "The Application of Binary Paired Comparisons to Listening Tests," proceedings of a Symposium on Perception of Reproduced Sound, Bech, S. and Pedersen, O.J. (eds.), Aug. 30–Sept. 2, Gammel Avernaes, Denmark, pp.71-80 (1987).

[50] ISO, ISO 8253-1, "Acoustics - Audiometry methods - Part 1: Basic pure tone air and bone conduction threshold audiometry," International Organization for Standards (1989).

[51] ISO, ISO 389 "Acoustics- Reference zero for the calibration of audiometric equipment - Part 1: Reference equivalent threshold sound pressure levels for pure tones and supra-aural earphones," International Organization for Standards (1998).

[52] S. Olive, "A New Listener Training Software Application" preprint 5384, presented at the 110th Audio Eng. Soc. Convention (April 2001).

[53] F.E. Toole, "Loudspeaker Measurements and their Relationship to Listener Preferences: Part 1", J. Audio Eng. Soc., vol. 34, pp. 227-235 (1986).

[54] F. E. Toole, "Loudspeaker Measurements and Their Relationship to Listener Preferences," J. Audio Eng. Soc., part 2, pp.323-348 (1986 May).

[55] FE. Toole and S.E Olive, "Hearing is Believing vs. Believing is Hearing: Blind vs. Sighted Listening Tests, and Other Interesting Things," Preprint Number: 3894 Convention: 97 (October 1994)

[56] A. Gabrielsson, U. Rosenberg and H. Sjogren, "Judgments and dimension analyses of perceived sound quality of sound-reproducing systems," J. Acoust. Soc. Am., vol. 55, no. 4, pp. 854-861 (1974).

[57] A. Gabrielsson. "Dimension analysis of perceived quality of sound reproduction systems," Scand. J. Psychol. vol. 20, pp. 159–169 (1979).

[58] A. Gabrielsson, B. Hagerman, T. Bech-Kristensen and G. Lundberg, "Perceived Sound Quality of Reproductions with Different Frequency Responses and Sound Levels," J.

Acoust. Soc. Am., vol. 88, pp. 1359-1366 (1990).

[59] S. Bech, "Selection and Training of Subjects For Listening Tests on Sound-Reproducing Equipment," J. Audio Eng. Soc., vol. 40, pp. 590-610 (July/Aug 1992).

[60] S. Olive, "Differences in Performance and Preference of Trained versus Untrained Listeners in Loudspeaker Tests: A Case Study," J. Audio Eng. Soc., vol. 51, no. 9, pp. 806-825 (September 2003).

[61] Y. Ando, "Subjective Preference in Relation to Objective Parameters of Music Sound Fields with a Single Echo," J. Acoust. Soc. Am., vol. 62, pp. 1436–1441 (1977).

[62] F. Rumsey, S. Zielinski, R. Kassier, and S. Bech,"On the relative importance of spatial and timbral fidelities in judgments of degraded multichannel audio quality," J. Acoust. Soc. Am vol. 118, no. 2, pp. 968-977 (August 2005).

[63] F. Rumsey, "Controlled Subjective Assessments of Two-to-Five-Channel Surround Sound Processing Algorithms," J. Audio Eng. Soc., vol. 47, no, 7/8, pp. 563-582 (July/August 1999).

[64] Toole, F.E. "Subjective Measurements of Loudspeakers: A Comparison of Stereo and Mono Listening," Preprint Number: 2023 Convention: 74 (September 1983)

[65] Personal communication with Floyd Toole (2007 August).

[66] R. C. Heyser, "Determination of Loudspeaker surfaces that absorb the mid and higher frequencies Signal Arrival Times, Parts I, II, and III," J. Audio Eng. Soc., vol. 19, pp. 734-743 (1971 Oct.).

[67] John Vanderkooy and Stanley P. Lipshitz," Uses and Abuses of the Energy-Time Curve," J. Audio Eng. Soc., vol. 38, no. 11, pp. 819-836 (November 1990).

[68] G.A. Soulodre, M.C. Lavoie and S.G. Norcross, "Objective Measures of Listener Envelopment in Multichannel Surround Systems," J. Audio Eng. Sp., vol. 51, no. 9, pp. 826-840 (September 2003).

[69] J.S. Bradley and G.A. Soulodre, "Listener Envelopment: An Essential part of Good Concert Hall Acoustics, " J. Acoust. Soc. Am., vol 99, p.22 (1996 Jan.).

[70] J.S. Bradley and G. A. Soulodre, "The influence of late arriving energy on spatial impression," J. Acoust. Soc. Am., vol. 97 (4), pp. 2263-2271, (1995).

[71] G. A. Soulodre, N. Popplewell and J.S. Bradley, "Combined Effects of Early Reflections and Background Noise on Speech Intelligibility," J. Sound Vib. vol. 135, pp. 123–133 (1989).

[72] T. Okano, L.L. Beranek and T. Hidaka "Relations among interaural cross-correlation coefficients (IACCE), lateral fraction (LFE), and apparent source width (ASW) in concert halls", J. Acoust. Soc. Am. vol. 104, no. 1, pp. 255-265, (July 1998).

[73] S.E. Olive and F.E. Toole," The Detection of Reflections in Rooms," J. Audio Eng. Soc., vol. 37, no. 7, pp. 539-553 (July/August 1989).

[74] Ernst-Joachim Voelker, "Control Room for Music Monitoring," J. Audio Eng. Soc., vol. 33, no. 6, pp. 452-462 (June 1985).

[75] F.E. Toole and S. Olive, "Perception and Measurement of Resonances", J. Audio Eng. Soc., vol. 36, no. 3, pp. 122-141 (1989).

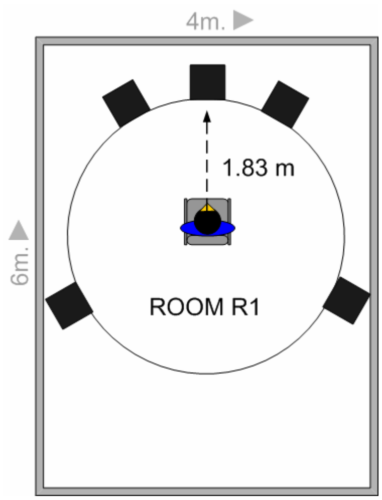
[76] A. Weisser and J.H. Rindel, "Evaluation of Sound Quality, Boominess, and Boxiness in Small Rooms," J. Audio Eng. Soc., vol. 54, no. 6, pp. 495-511 (June 2006)

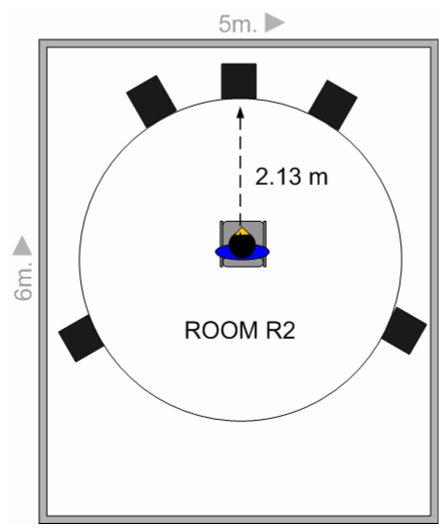
[77] A. Devantier, "Characterizing the Amplitude Response of Loudspeaker Systems," presented at the 113th Convention of the Audio Engineering Society, J. Audio Eng. Soc. (Abstracts), vol. 50, p. 954 convention paper 5638 (Nov.2002).

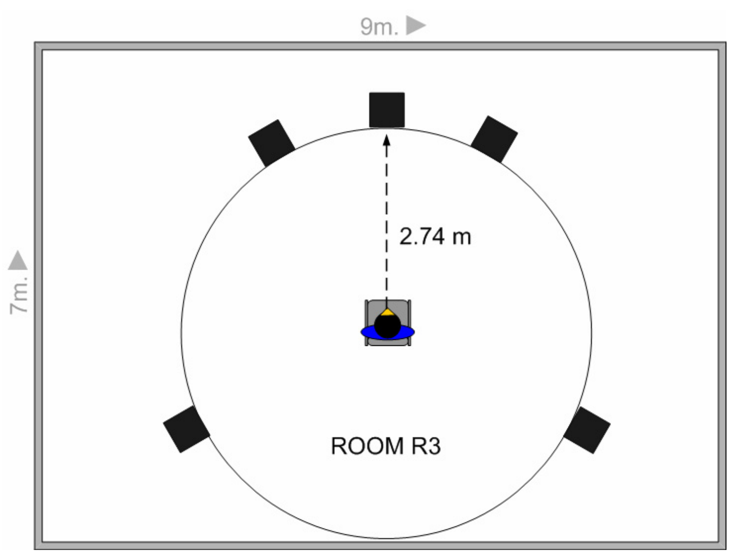
8. APPENDICES

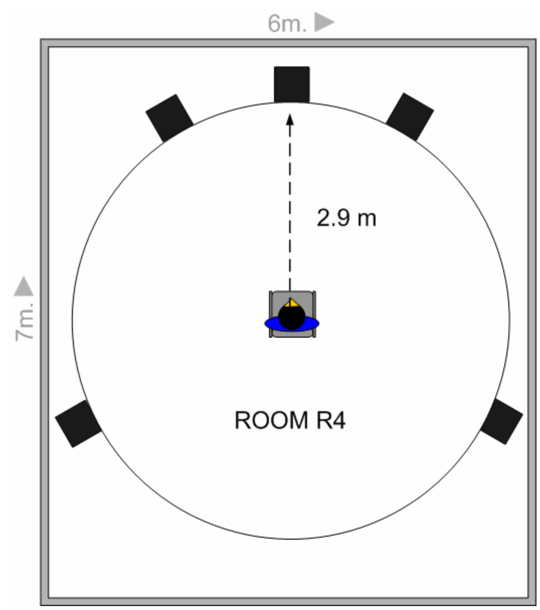
8.1. APPENDIX 1: Room and Loudspeaker Setups

Below are the diagrams of the loudspeakers in each listening room used in this study. Starting from the top left and going clockwise are rooms: R1, R2, R3, and R4. The listening distances from the loudspeakers are shown for each room as well. All setups are compliant with ITU-R BS. 775.1 [33] with the left and right front channels at ±30º, the center speaker at 0º, and the left and right surrounds at ±115º.



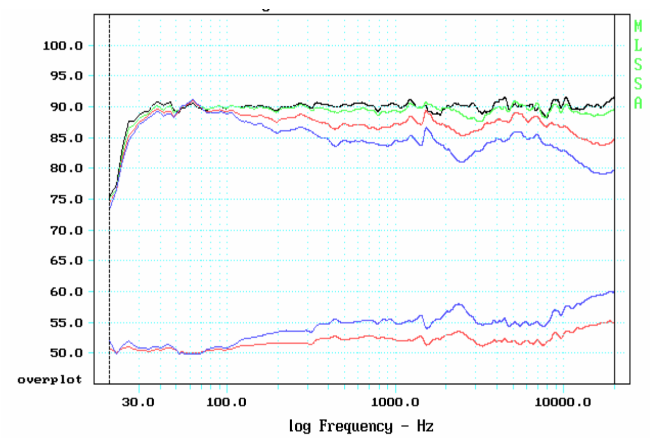




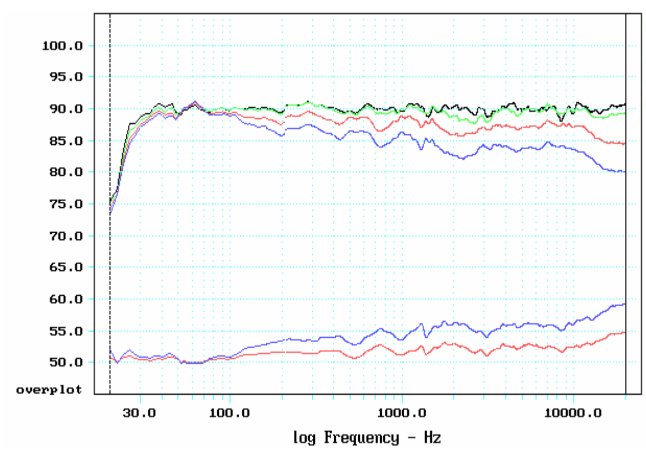


Appendix 2: Loudspeaker Measurements

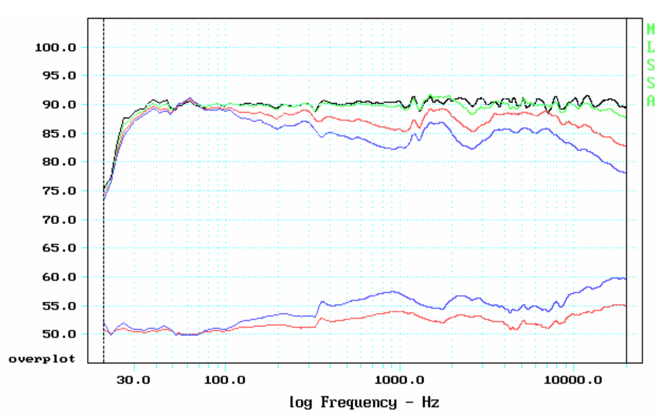
Below are the on-axis and spatially-averaged anechoic frequency response measurements of the four loudspeakers used in the experiments: (a) L1 (3-way with no waveguide on the tweeter), (b) L2 (3-way with waveguide) (c) L3 (2-midranges vertically arranged with the tweeter in the middle, and (d) L4 (same as L3 but the midranges are arranged horizontally) The curves in each graph represent from top to bottom: the on-axis response (black curve), the average direct sound (green curve), the early-reflections (red curve), and the sound power response (blue curve). The two bottom curves represent the directivity indices based on the sound power (blue curve) and the early reflections (red curve). For more details on how these measurements are derived see [77].



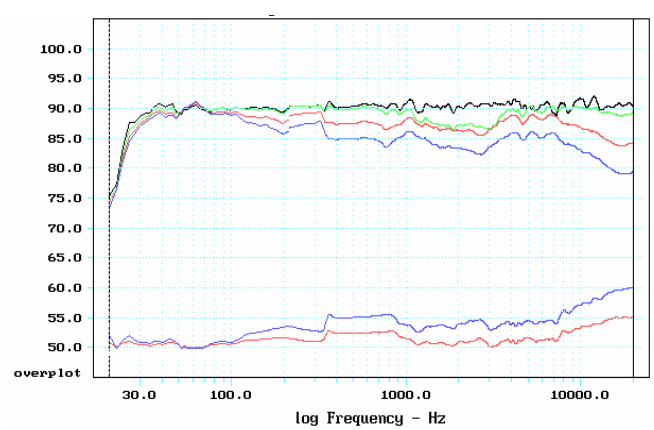
(a)



(b)



(c)



(d)

8.2. Appendix 3: Listener Instructions

In these tests you will be evaluating binaural headphone reproductions of different 5.1 channel surround sound systems. Your preferences among the different surround sound system should reflect your overall personal tastes and preferences in sound reproduction quality. In formulating your preference choices and ratings, consider both the spectral and spatial qualities of each surround sound system.

Paired Preference Choice Task

You are asked to first complete 6 trials, where you are presented pairs of different surround systems. You switch between the two surround sound systems by clicking on Buttons A and B using your stylus (see Fig. 1) and listen as long as you like. Enter a rating of 1 for the system you prefer, and a rating of 0 for the system that is not preferred. The rating can be changed by moving the slider with your stylus or using the Up-Down buttons on the PDA. Once your final preference choices are made click on the Buttons OK, and then DONE (note: If you give tied ratings, you will be prompted to re-enter your ratings).

After you have hit DONE, your results will be automatically saved and the next trial will be loaded. The same program selection (Jazz Vocal) is used for all 6 trials. When you have completed all 6 trials, the experimenter will instruct you to begin the preference rating test.

Preference Rating Task

In this test, you will complete 9 trials where four different surround sound systems are presented in each trial. Three different program selections will be used in this test. Like the previous test, you can listen and switch among the four different surround systems as long as is necessary by clicking on Buttons A through D (see Fig. 2). Using the slider and/or Up-Down PDA buttons, enter a rating for each surround system using the 11-point preference shown below.

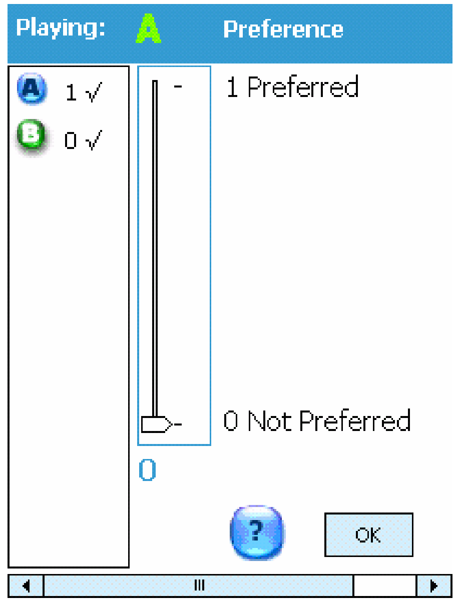


Figure 1 The graphical user interface for the preference choice test.

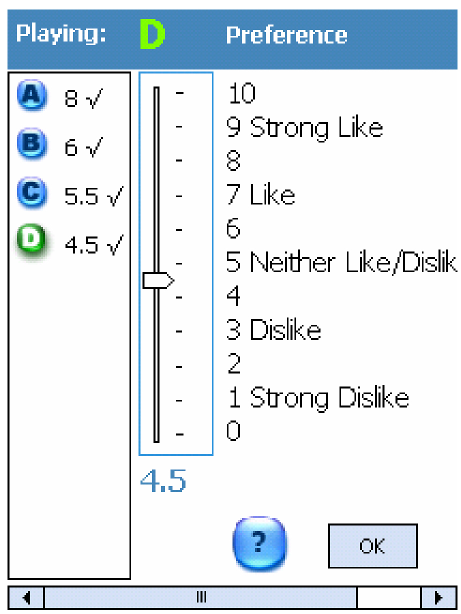
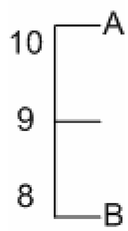


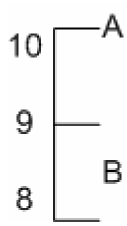
Figure 2 The graphical user interface for the preference rating test.

We are more interested in the relative differences in preferences among the different surround sound systems, rather than how you feel about them on an absolute basis of overall sound quality. Therefore, we encourage you to use as much of the scale as necessary in order to discriminate and express your relative preferences among each surround sound system presented. We give the following guidelines in how you express the relative differences in preference among two different surround sound systems.

A strong preference between two surround systems (e.g. A and B) is expressed as separation of 2 or more ratings.



A moderate preference is a separation of 1 to 2 ratings.



A slight preference is a separation of 0.1 to 0.5 rating.