Subjective Methodology and Results of AM Nighttime Transmission Testing

July 1, 2003

Prepared for iBiquity Digital Corporation by:

Ellyn Sheffield, PhD
Sheffield Audio Consulting
10 Alta Vista Drive
Princeton, New Jersey 08540
(609) 737-9451
egsheffield@comcast.net

Subjective Test Methodology

This subjective test methodology evaluated the impact of IBOC on AM analog nighttime transmission. In order to realistically assess consumer reaction to nighttime transmissions (with and without IBOC), a new Absolute Category Rating (ACR) 5-point scale was developed. Designing this new scale was necessary for two reasons. First, AM nighttime listeners do not fit the profile of typical radio consumers in that they are highly motivated to listen to the program that they have selected, and will endure worse transmission conditions than casual radio consumers. Because they are highly motivated to listen, the importance of overall quality (i.e. attractiveness of sound) is less likely to impact their behavior than other factors, such as intelligibility, annoying interference, sustained loss of signal, etc. Second, the quality of typical AM nighttime transmission would likely be considered either "fair" or worse by the average consumer and therefore the standard quality ratings (i.e. Excellent, Good, Fair, Poor, and Bad) would be overly compressed, providing little insight into the differences between IBOC-off and IBOC-on transmissions. For these reasons, the adapted scale asked participants to rate samples on the basis of whether they would continue to listen to the program or switch to another, depending on how motivated they were when listening to the transmission. Results from this question provided relational information, consistent with typical ITU recommended MOS scales, but also provided **threshold** information which exposed the point at which consumers would no longer listen to the broadcast.

1.1 Adapted 5-point scale

Table 1.1 describes the scale that participants used to judge all transmissions. Notice that threshold information is obtained at two places, "3.0" (i.e., the participant would continue to listen *only when they were motivated*) and "1.0" (the participant would always turn off the broadcast).

Table 1.1: Adapted 5-point scale

Rating	Description of Rating (as provided to test subjects)	Numeric Translation for Analysis
Unimpaired (Keep On)	This sample sounded good. I would listen to this audio under all circumstances.	5.0
Somewhat Impaired (Keep On)	This sample sounded good, but I heard some background impairments and noise. Still, I would listen to this audio a majority of the time.	4.0
Noticeably Impaired (Keep on if Motivated)	This sample was intelligible, but the background chatter and noise was noticeable and significant. I would continue to listen to this audio a majority of the time only if I was extremely interested in the program.	3.0
Severely impaired (Keep on only sometimes if extremely motivated)	This sample was mostly intelligible but the background chatter and noise was very annoying. I would continue to listen some of the time only if I was extremely interested in the program.	2.0
Failed (Turn off)	This sample is unintelligible. I would not listen to this audio under any circumstance.	1.0

1.2 Audio samples

Due to the nature of nighttime transmission, samples at the same or similar D/U signal levels varied widely in both overall quality and size of impairment. This variation made it difficult to characterize an entire listening experience based only on one sample-pair[1]. Therefore, where possible, the test included several sample-pairs at the same D/U level. The number of sample-pairs chosen to be included at each D/U levels was based on the total number of samples collected during field recordings divided by the number of samples recorded at that D/U level. Thus, each D/U level listed in "Table 2.1 - Experimental Conditions", contains at least 1 on-off sample pair, and potentially contains up to 4 sample-pairs, depending on how many recordings were made in the field at that D/U level, and how many samples met the sample selection criteria (see section 1.2.2 for details on selection of samples).

In total, 262[2] audio samples were presented to participants for rating, 248 field recordings (from 124 sample-pairs), and 14 laboratory generated samples. Field samples included Sky-to Ground, Ground-to-Sky and Sky-to-Sky transmissions between -10 and +10 D/U. Samples were either considered "on-axis" if there were on or near the direct line between the two stations (WLW and WOR) or "off-axis" if they were not on a direct line between the two stations. High quality, unimpaired laboratory-generated samples were included to provide participants the opportunity to hear transmissions they would rate highly (i.e., 4 or 5). They were also intended to help alleviate the monotony resulting from presenting the same field samples to participants multiple times. All participants heard both IBOC-Off and IBOC-On samples for all conditions.

Transmissions recorded over three receivers were included in this test. These include the Delphi, Sony, and GE Superadio receivers. Table 1.2 lists the receiver, model number and type.

Table 1.2: Description of Receivers

Manufacturer	Model Number	Type
Delphi	09394139	Auto
Sony	CFD-S22	Boom-Box
GE	7-2887A	Portable

1.2.1 Receiver Selection Criteria

The Delphi was chosen since it is widely available, has excellent front-end performance and has narrowband filtering. The Sony was chosen to represent the semi-portable "boombox" class of receivers. It is battery-powered and is often used outdoors. Both the Delphi and Sony receivers were included in prior AM and FM interference tests. The GE Superadio receiver was selected because it represents the higher end of the portable receiver market. It also can be powered by

[1] A sample-pair consists of two samples taken from a one-minute recording segment, one being IBOC-on, the other being IBOC-off.

[2] Not all 262 recordings were used in data analysis. Some sample-pairs were omitted because in post-test analysis it was found that the D/U averages for "on" and "off" samples in specific sample-pairs did not match closely enough for comparisons to be made. See Section 4 for a complete discussion on this point.

batteries and taken outdoors. It has a large internal antenna that can pick up distant signals. Thus, all of these receivers are used outdoors by consumers, away from the man-made noise typically generated in office buildings and homes.

Two other receivers were used during the field testing: the Technics home hi-fi receiver and the Pioneer auto receiver. Neither was used for this audio test[3].

1.2.2 Processing Audio Samples (Recording, Selecting, Editing and Leveling)

All field samples were collected under NRSC auspices during August and December, 2002. As with the iBiquity daytime AM field test program, audio was recorded at 30-second intervals, alternating between IBOC-Off and IBOC-On. For each test condition in this study, 2 samples were chosen from "on-off" or "off-on" 60-second segments. All individual sound samples were edited, labeled and leveled for presentation to participants. Resulting samples were 6 to 10 seconds long. Only sample-pairs (IBOC-off, and IBOC-on) that were matched in genre, density and programmatic material were included in this test plan.

Specifically, samples included in this plan were based on the following criteria:

- the IBOC-off and IBOC-on samples were matched for genre (i.e., speech to speech; voiceover to voiceover; commercial to commercial)
- the programmatic content was appropriate (e.g., programming will be included only if it is considered neutral and non-offensive)
- talkers' intelligibility was consistent and clear (heavily accented speech, garbled speech and stuttering was minimized)
- the speech density was equivalent between IBOC-off and IBOC-on samples (i.e., no long pauses in speech for one sample but not the other)
- within a given condition, if there were different announcers for the IBOC-off and IBOC-on samples, the announcers' voices were matched vis-à-vis pitch and rate of speech.

[3] The Technics relies on AC current and is rarely used outdoors. Its performance is severally impacted by background noise generated from TVs, computers, fluorescent lights etc, which are typically found in home and office environments. The signal reaching the antenna is degraded because many homes are sided with aluminum. Therefore, it is not a candidate for nighttime listening in most parts of the country. The Pioneer auto receiver has slightly wider front-end filtering than the Delphi, therefore it is not as good at receiving stations at night as the Delphi. It was eliminated as a test receiver since the differences between IBOC-on and IBOC-off would not be as obvious. Therefore, the Delphi represents a more conservative choice than the Pioneer for this test program.

1.3 Experimental Design

Table 1.3 shows the experimental design of the study. Notice that there varying numbers of sound samples at different D/U levels. Because it was difficult to find multiple sample-pairs at consistent D/U levels that also matched all of the qualifications described in Section 1.2.2, on-off sample-pairs were included at a D/U level if the midpoint between the D/U level for the "on" and "off" samples was within ±2.5 dB of that level[4]. For example, if the average of an "off" sample in a sample-pair was 1.76dB and the average of its corresponding "on" sample was -3.04dB, their span would be 1.76 + 3.04 or 4.8dB, and their midpoint would be 1.76dB - 2.4dB = -.64dB. Since -.64dB is within ±2.5 dB of 0, the pair would be placed in the +0 D/U category. The averages for all samples by sample-pair are included in **Appendices C, D and E.** Although 262 samples were played for participants, only 218 are included in the Experimental Design (see Section 4 for a description of post-hoc D/U analysis of sound samples for a full explanation).

Table 1.3: Experimental Conditions

Condition	D/U Range	Delphi		SONY		GE		Total
		Off	On	Off	On	Off	On	
OFF AXIS								
Sky-to-sky	-10	2	2	2	2	2	2	12
	-5	0	0	0	0	0	0	0
	0	2	2	2	2	2	2	12
	+5	2	2	2	2	2	2	12
	+10	2	2	2	2	2	2	12
Sky-to-ground	-5	1	1	1	1	1	1	6
	0	1	1	1	1	1	1	6
	+5	1	1	1	1	1	1	6
	+10	1	1	1	1	1	1	6
Ground-to-Sky	-5	1	1	1	1	1	1	6
	+0	1	1	1	1	1	1	6
ON AXIS								
Ground-to-Sky	-10	2	2	2	2	2	2	12
	-5	4	4	4	4	4	4	24
	+0	3	3	3	3	3	3	18
	+5	0	0	0	0	0	0	0
	+10	1	1	1	1	1	1	6
Sky-to-Ground	-5	1	1	1	1	1	1	6
	+0	2	2	2	2	2	2	12
	+5	4	4	4	4	4	4	24
	+10	3	3	3	3	3	3	18
Laboratory Samples								14
TOTAL SAMPLES								**218**

[4] Spectral Data was collected approximately every second, and each plot was a rolling average of the previous 10 samples. A figure representing the averaged D/U for each audio test file was obtained by converting all D/U dB measurements taken during the sample to voltage, averaging them and converting that figure back to dB.

1.4 Participants

46 subjects (24 males and 22 females) participated, distributed between 16 and 65 years of age. Forty-three participants were from the general public, 3 participants were representatives from the NAB Ad-Hoc Technical Group. One participant was excluded because she did not finish the test due to computer problems. Two participants were excluded because post-hoc statistical analysis indicated that their pattern of ratings were significantly different from the patterns of ratings for the whole group. In order to demographically characterize the test sample, participants were asked to fill out a questionnaire prior to taking the test, which included their age, gender, and whether they listen to AM radio on a regular basis (see **Appendix A** for a sample questionnaire). Listeners who reported that they listened to AM daily were classified as "Heavy AM listeners"; those who reported that they listened to AM within the last week or month were classified as "Light AM listeners" and those who reported that they listened within the last year or not at all were classified as "No AM listeners". See Tables 1.4-1 and 1.4-2 for the demographic breakdown. Participants who claimed hearing loss due to temporary or chronic problems were excluded from participating.

Table 1.4-1: Participants age and gender

Age	Male	Female
18-29	5	5
30-39	5	6
40-49	5	5
50+	7	5

Table 1.4-2: Number of participants listening to AM

Age	Heavy AM listening		Light AM listening		No AM Listening	
	Male	Female	Male	Female	Male	Female
18-29	1	2	0	3	2	2
30-39	3	2	1	2	1	2
40-49	4	1	0	1	1	3
50+	4	1	2	1	1	3

1.5 Procedure

Participants listened to each sample once. They were encouraged to listen to the sample again if they needed more time to rate it. The order of sample presentation was randomized for each participant. Participants listened to 67 or 66 trials in a listening session, followed by a 5-minute break. The total time for an experiment, including training, testing and breaks was approximately 100 minutes.

Sound samples were presented to participants over loudspeakers in acoustically appropriate test environments (see Section 2.0 for details).

1.6 Training Period and Screening

Training included an orientation to the software used to collect data and a description of the scenarios on which participants based their answers. Experimenters described each category in the 5-point scale at length, ensuring the participants understood the difference between listening in a "casual" way and listening with "motivation" (see **Appendix B** for details). Participants were directed to rate samples based on the quality of the transmission and not the programmatic material. Participants were also told that these were AM recordings taken from real radio programs from around the country. In order to minimize the risk of biasing participants, training samples were not played prior to testing, nor was information given about impairments that they might hear. Screening was performed after data collection took place. A post-hoc statistical test was conducted for each participant to ensure that individual participants' vector of rating correlated positively to the vector of ratings found in the group. If a person's pattern of answers were significantly different from the group's, their data was not included in analyses.

2 Audio Playback Setup and Testing Environment

Participants were tested individually using iBiquity software. All audio samples were presented to listeners over loudspeakers. A set of medium quality auto loudspeakers, the Optimus (Tandy – Cat. #12-1773) were used to deliver the audio samples to participants. All manufacturer's suggestions for requirements for optimal performance were followed, including amplification.

Since loudspeakers were used for testing, it was important for the test environment to be quiet, free from aural and visual distractions. Listening rooms were configured for testing with low ambient/background noise and minimal ingress of external sounds. Ambient/background noise did not exceed 43dB(A)[5]. Each test participant was located in a pre-determined position within the room, and was instructed not to move or relocate the chair during the course of the experiment. Loudspeakers were configured in the room for optimal listening performance. Figure 2.1 shows an example of the test set-up.

Figure 2.1: Experimental room 1

[5] Measurements were taken with a TerreSonde Audio Toolbox, A-weighted in slow response mode.

3 Long-sample test

After completing the short sample test, a subset of participants was asked to rate 24 additional samples that were between 24 and 28 seconds in length. This test was designed to examine whether consumers would rate long samples differently than 6-10 second, short samples. It is believed that certain impairments cause "listener fatigue" and that over longer periods of time participants become more critical of samples due to increased exposure to these impairments. Would this be true for AM nighttime transmissions, and would the introduction of IBOC exacerbate this effect? By creating short and long samples from the same source material and presenting them to participants, it was possible to determine whether consumers would judge transmissions more critically merely due to the length of the presentation sample. Long sound samples were parsed and edited identically to short ones except that they lasted approximately four times longer. Thus, for a given condition (e.g., Sky-to-Ground +10 D/U), a long sample included the short sample and 16-18 additional, contiguous seconds taken from the original 30-second transmission segment. Five conditions were tested: Ground-to-Sky -10 and -5 D/U; Sky-to Ground +10 D/U; and Sky-to-Sky +0 and +10.

4 Post-hoc analysis of samples

As was noted in Section 1.2.2. each sample-pair was chosen from 60 continuous seconds of field transmission, resulting in two individual samples (an IBOC-off sample, taken from the first 30 seconds and an IBOC-on sample, taken from the next 30 seconds[6]) The D/U dB level for each resulting sample-pair was taken from a reading of the field intensity at the top of the minute. However, because we presented to participants only 6-10 seconds of the original 30-second recording for each sound sample, it was critical to re-check the average D/U dB level for each shortened sample. Upon re-calculating the dB level for individual samples, we found several cases in which the difference between IBOC-on and IBOC-off was too large for meaningful comparisons to be drawn. Thus, we eliminated those samples where the absolute difference in average dB between IBOC-on and IBOC-off was greater than 7dB.

5 Results

5.1 Interpreting participants results using the 5-point rating scale

When interpreting participants' ratings, it is important to keep in mind that the rating scale used for this study does not follow the same principles as does the ACR-MOS quality rating scale. The ACR-MOS rating scale asks participants to focus solely on one dimension – sound quality – while making their decision. The categories (Excellent, Good, Fair, Poor and Bad) were designed to be evenly spaced. The rating scale used in this study asks participants to focus on two dimensions simultaneously: (a) the extent to which impairments are heard (5 = No impairments heard; 4 = Impairments heard, but not bothersome; 3 = Significant impairments heard; 2 = Significant, disruptive impairments; 1 = complete failure), and (b) whether they would continue to listen to the sound sample depending on their perceived motivation (a rating of 4 or 5 signifies that the participant would listen all the time; 3 – participants would listen only if

[6] In point of fact, some 60-second recordings were actually 30 seconds IBOC-on followed by 30 seconds IBOC-off; others were 30-sec. IBOC-off followed by 30 seconds IBOC-on.

motivated; 2 – participants would listen rarely even if motivated; 1 – participants would never listen). The categories were not chosen to be evenly spaced. They are, instead, distinct decision points. Participants must chose between them considering their "state of mind" as well as the level of impairment heard in the sample. Using a numeric translation of this categorical scale, an individual score of 4.0 (Somewhat impaired – keep on) or 5.0 (Unimpaired – keep on) indicates that a participant would listen to the transmission all of the time, regardless of impairments heard or their level of motivation. 3.0 is a particularly interesting demarcation, because at this score participants claim they would listen a majority of the time *if they were motivated* (which best characterizes nighttime listeners), yet they still claim they hear noticeable impairments. Conversely, at 2.0 participants claim that the sample is severely impaired and that they would listen only very infrequently, when they were extremely motivated. Therefore, somewhere between 3.0 and 2.0 there is critical point at which a *majority of listeners* would no longer choose to listen, even when motivated to do so. In order to determine this point, participants' scores were re-coded using the following conversion: if a participant rated a sample as a 1 (Failure) or a 2 (Would listen only under extraordinary circumstances) , it was re-coded to "0" (meaning that they would almost always turn the broadcast off); if a sample had received a 3, 4 or 5, it was re-coded as a "1" (meaning that they would continue to listen to the broadcast). See Table 5.1 for conversions. The resulting "on-off score" for each sound sample was simply the proportion of participants who would continue to listen to it. The original mean scores were compared to these "on-off scores" to determine the point at which the majority of participants would continue to listen.

Using this translation, at the 3.0 level, approximately 68% of all listeners would keep listening to the sample. Notice that because these are aggregated scores, 100% agreement that the transmission is acceptable (when participants are motivated) is not realized at 3.0. In fact, 100% agreement occurs only at approximately 3.8. At the 2.6 level, approximately 50% of listeners would still keep the radio on. Below 2.6, a majority of listeners claim that they would turn the program off. Thus, 2.6 is a significant cut-off point, as it reflects when the majority of people would still be satisfied with transmission quality, if they were motivated to listen to the program.

Table 5.1: Conversion from mean opinion score to on/off rating

Original Mean Score	On/Off Conversion	Meaning
1 and 2	0	Would not listen
3, 4, and 5	1	Would listen

5.2 Preliminary Analyses

In order to determine whether participants reacted differently to samples because of their age, gender, and experimental room in which they were tested, preliminary analyses of variance (ANOVAs) were conducted. A 2 (Gender) x 2 (IBOC: on/off) ANOVA was performed on participants' ratings. Although this analysis showed a significant effect of IBOC (i.e., IBOC-off was rated significantly higher than IBOC-on) it showed no effect of gender. Thus, females and males rated samples similarly, regardless of whether IBOC was off or on. A 4 (Age) x 2 (IBOC: on/off) analysis of variance was additionally performed. This analysis showed a significant effect of age, although the differences were minor. With the exception of 18-29 year olds, the older the participants were, the more tolerant they were, rating samples significantly higher. However, in this study, 18-29 year-olds were also quite tolerant, rating samples higher than 50-

59 year olds. While interesting, this finding is not corroborated by other tests which have found that younger listeners are generally more critical than older listeners. Again, IBOC did not interact with age, suggesting that IBOC did not play a mediating role in participants' judgments. Finally, a 2 (room A; room B) by 2 (IBOC: on/off) analysis of variance was performed to see whether scores might be affected by the different environments in which people were tested. The average score of all participants combined for all samples for Room 2 was **2.8**; the average for Room 1 was **2.9**. Although these averages are statistically different, the difference is minimal. Table 5.2 shows the means for gender, age, and room placement.

Table 5.2 Participant Ratings by gender, age and room placement

	Female		**Male**	
	Room 1	**Room 2**	**Room 1**	**Room 2**
18-29	2.93	2.79	2.81	3.13
30-39	2.58	2.95	2.71	2.42
40-49	3.29	2.64	2.64	2.80
50+	3.30	2.85	2.85	2.69

5.3 Reference samples (high anchors)

Recall that 14, laboratory generated, unimpaired references were included in this test to ensure that participants heard recordings that they would rate a "5". The mean score of all rated reference samples was 4.84, indicating that participants were well "calibrated" during the testing procedure, that they could easily distinguish impaired and unimpaired samples, and that they were willing to appropriately use all 5 categories in the 5-point scale.

5.4 AM listeners vs. Non-AM listeners

Because this subjective study was designed to evaluate customer satisfaction for AM nighttime transmissions, it was important to evaluate test data in relation to participants' day-to-day listening habits, in order to determine whether listeners who listened to AM regularly would rate transmissions differently from those listeners who claimed they did not listen to AM regularly. It was hypothesized that because regular AM listeners were more familiar with the "AM sound", they would be more likely to have a realistic internal representation of it and, therefore, rate the sound samples more favorably than those listeners with no AM experience. To test this hypothesis, a 3 (Heavy AM; Light AM; No AM) x 2 (IBOC: on/off) ANOVA was conducted on rating scores. There was a significant main effect of listeners, but this difference did not interact with IBOC, indicating that the introduction of IBOC did not have a negative effect on any particular group of listeners. Listeners claiming *more* experience with AM (the Heavy AM and Light AM groups) rated samples significantly *lower* than listeners with no experience. See Table 5.4 for details. This finding was somewhat surprising. Intuitively, it seemed likely that listeners who were exposed to AM on a regular basis would have rated it higher than listeners who had not been exposed to AM. This was not the case. Nevertheless, this finding may be fortunate – because our sample population contained a large number of AM listeners, results are conservative, and will most likely accurately depict the real-world listening experience.

Table 5.4: Mean scores from "Heavy", "Light" or "No AM-listeners

AM Listener	IBOC OFF	IBOC ON
Heavy	3.20	2.94
Light	3.19	2.91
No	3.43	3.21

5.5 Short vs. Long Samples

In order to test whether participants judged short (6-10 second) samples differently than longer (24-26 second) samples, a 2 (Short vs. long) x 3 (AM Heavy, AM light and No AM listener) x 2 (IBOC-on; IBOC-off) ANOVA was conducted. This analysis not only considered whether participants would become more critical over time, but also examined whether there was any difference in the way participants perceived IBOC over time. Additionally, this analysis looked at AM vs. non-AM listeners. There was no effect of short vs. long samples, and no interactions with IBOC or AM vs. non-AM listeners. (See Table 5.5 for a comparison of mean scores.) Therefore, with minor exception[7] whether listeners heard short samples or longer samples, they rated the samples similarly. This is another surprising finding because there is a great deal of speculation that people grow more dissatisfied with impaired audio transmissions as they listen for longer periods of time. It is possible that 24-26 seconds was not a long enough time for this effect to take place. However, when participants were debriefed after this test, no one suggested that the samples were too short to judge appropriately – they claimed that the samples were either too long, or approximately the right length. In fact, there is excellent reason to believe that people actually make decisions about samples within the first 6-10 seconds. Another de-brief question asked participants to think about when they made their final decision concerning their rating. Of the 38 participants, 12 stated they made their decision in the first 3 seconds; 24 within the first 6 seconds, and 2 within the first 10 seconds. Thus, it appears that the reason long and short samples were judged so similarly is a result of people's decision making strategy – listen briefly and decide to continue to listen or to turn the radio off.

Table 5.5: Comparison of Short and Long Samples

	Ground to Sky				Sky to Ground				Sky to Sky	
	Delphi Cut 4	GE Cut 1	GE Cut 4	SONY Cut 1	Delphi Cut 2	Delphi Cut 5	SONY Cut 2	SONY Cut 5	Delphi Cut 3	GE Cut 3
	D/U-5	D/U -10	D/U -5	D/U-10	D/U 10	D/U 10	D/U 10	D/U 10	D/U 10	D/U 10
IBOC OFF										
LONG	2.8*	2.8	2.6	1.0	4.1	4.6	3.3	3.8	4.2*	3.4*
SHORT	3.1	2.9	2.6	1.1	4.0	4.3	3.2	3.6	4.6	3.8
IBOC ON										
LONG	2.8	1.8	2.7	1.1	3.3	4.1	3.2	3.6	4.3	3.6
SHORT	2.6	1.8	2.7	1.1	3.3	4.2	3.0	3.6	4.3	3.5

[7] In 4 cases, short and long samples were rated differently (greater than .3). These occurred only in IBOC-off conditions. In 3 of these cases, participants rated longer samples lower than shorter samples, suggesting that some of the IBOC-off ratings may be slightly inflated. This anomaly suggests that people are more quickly able to judge the effects of IBOC impairments than when the signal is purely analog.

5.6 Effects of IBOC

Figures 5.6-1, 5.6-2 and 5.6-3 graphically depict the effect of IBOC on the analog transmission. Transmissions were placed into 3 groups, depending on their signal strength: (a) "strong interferer", including D/U ratios of -10 and -5dB; (b) "mid", including D/U ratios of +0 and +5dB, and (c) "weak interferer", or a D/U ratio of +10dB. The dotted line is the demarcation point: above the line, the majority of listeners would keep the program on. Below the line, the majority would turn it off. With the exception of Sky-to-Ground in strong interferer conditions (-10 and -5dB), the majority of motivated listeners would keep listening to their program after IBOC is introduced. Although ratings are generally lower for IBOC-on transmissions, they are nonetheless above or well above the 2.6 demarcation, indicating that the majority of participants would still be willing to listen to the transmission when motivated.

Figure 5.6-1: Ground-to-Sky

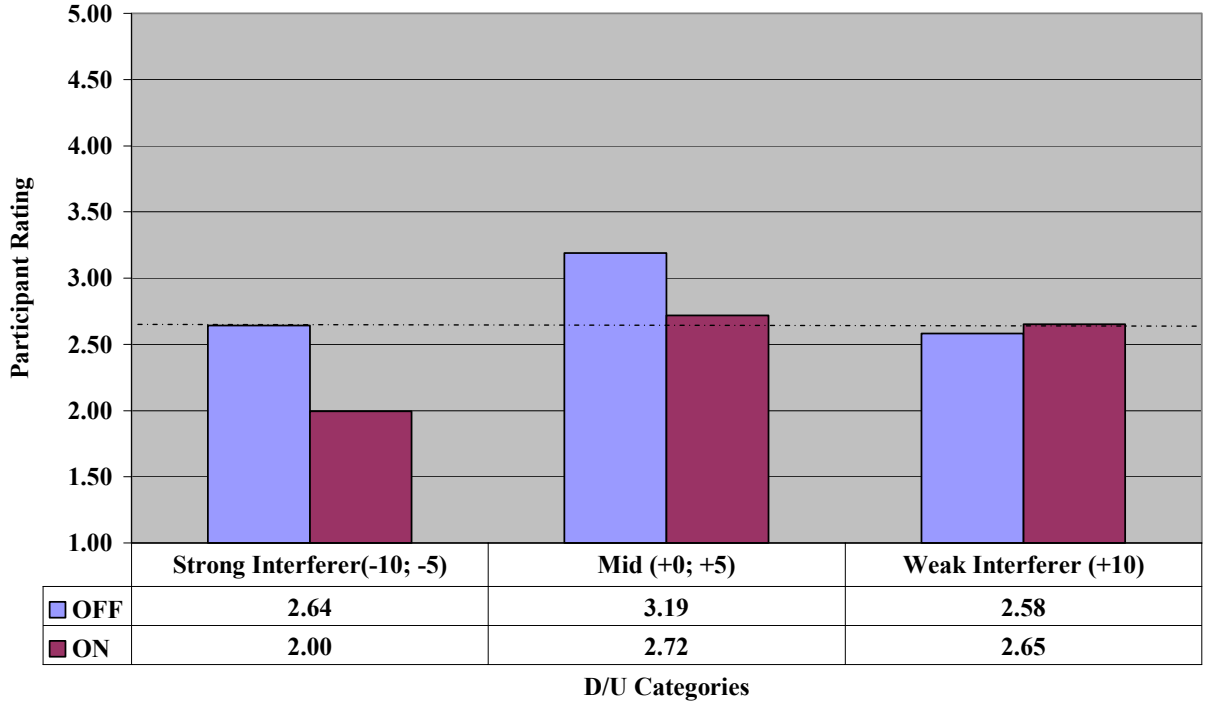


Figure 5.6-2: Sky to Ground

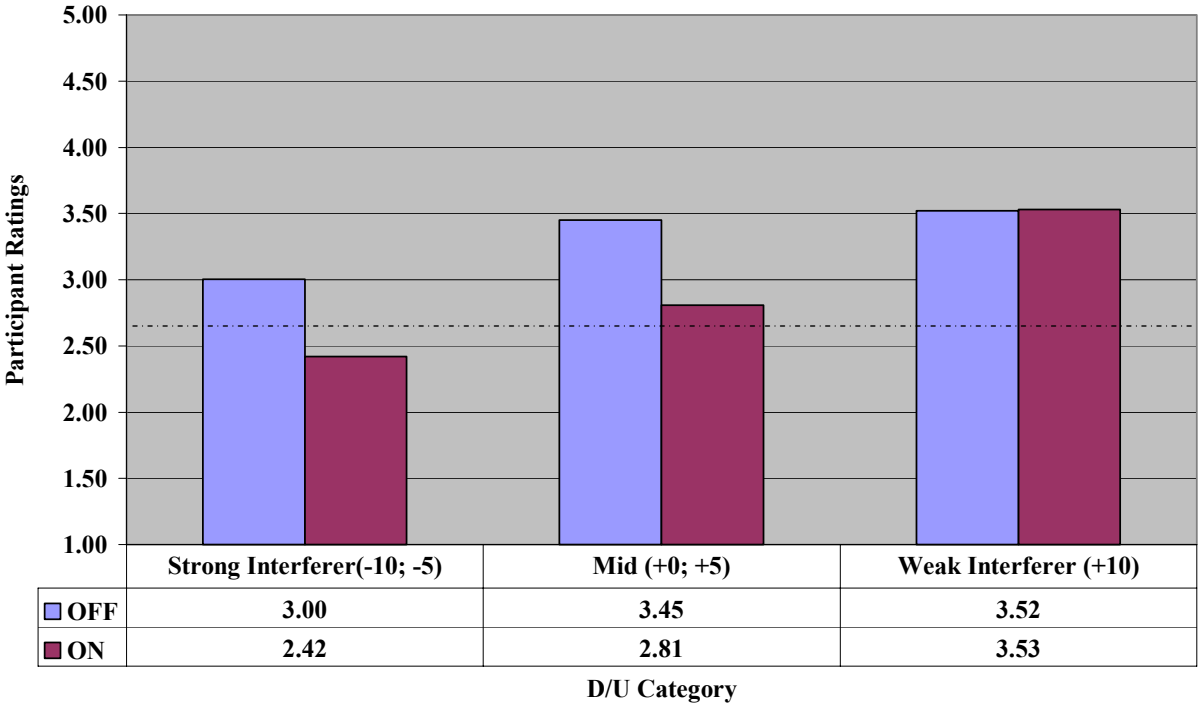
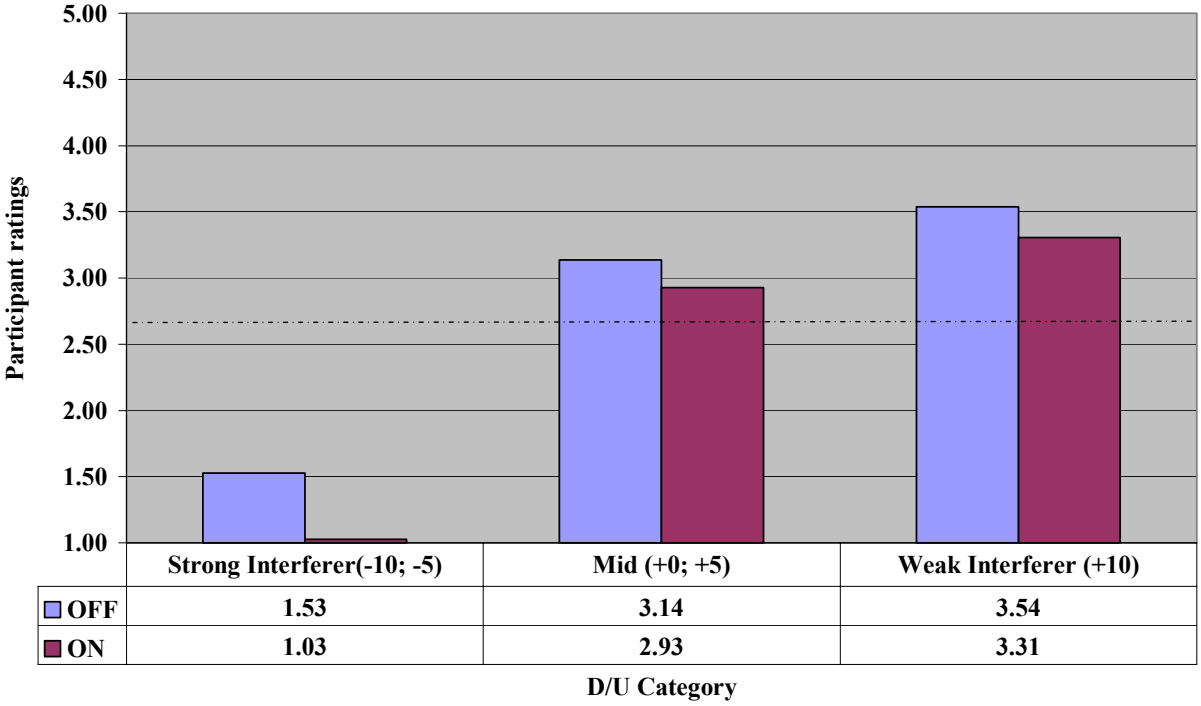


Figure 5.6-3: Sky-to-Sky



Tables 5.6-1 through 5.6-3 show rating scores and confidence intervals for samples aggregated by D/U level. Appendices C through E show individual sample rating scores, confidence intervals and the average D/U level of each 6-10 second sample.

Table 5.6-1: Mean Ratings by D/U: Ground to Sky

			Delphi		GE		SONY		TOTAL	
	D/U		OFF	ON	OFF	ON	OFF	ON	OFF	ON
OFF Axis	-5	Rating	3.7	2.8	3.5	3.1	3.6	3.1	3.6	3.0
		CI (+/-)	0.22	0.19	0.21	0.22	0.20	0.18	0.12	0.12
	0	Rating	3.6	2.7	3.5	3.0	3.6	2.9	3.6	2.9
		CI (+/-)	0.21	0.20	0.21	0.23	0.24	0.21	0.12	0.13
ON Axis	-10	Rating	2.3	1.9	2.3	1.5	1.0	1.0	1.9	1.5
		CI (+/-)	0.21	0.14	0.21	0.12	0.04	0.04	0.12	0.08
	-5	Rating	3.3	2.0	2.6	1.9	2.4	2.0	2.8	2.0
		CI (+/-)	0.14	0.12	0.15	0.13	0.18	0.12	0.10	0.07
	0	Rating	3.5	3.0	2.8	2.5	2.9	2.5	3.1	2.7
		CI (+/-)	0.15	0.11	0.19	0.13	0.17	0.16	0.10	0.08
	10	Rating	3.4	3.1	2.2	2.4	2.1	2.4	2.6	2.7
		CI (+/-)	0.21	0.25	0.22	0.22	0.20	0.19	0.16	0.14

Table 5.6-2: Mean Ratings by D/U: Sky to Ground

			Delphi		GE		SONY		TOTAL	
	D/U		OFF	ON	OFF	ON	OFF	ON	OFF	ON
OFF Axis	-5	Rating	1.9	1.7	3.1	3.4	3.0	2.9	2.7	2.7
		CI (+/-)	0.17	0.19	0.23	0.24	0.21	0.22	0.15	0.18
	0	Rating	2.7	3.4	3.6	2.6	3.4	2.6	3.2	2.9
		CI (+/-)	0.23	0.26	0.26	0.23	0.28	0.23	0.16	0.15
	5	Rating	3.8	2.9	3.6	3.3	3.7	3.5	3.7	3.2
		CI (+/-)	0.25	0.20	0.21	0.22	0.25	0.30	0.14	0.15
	10	Rating	3.4	3.9	3.3	3.6	3.1	3.5	3.3	3.7
		CI (+/-)	0.20	0.21	0.22	0.24	0.26	0.25	0.13	0.14
ON Axis	-5	Rating	3.5	2.5	3.0	1.5	3.6	2.5	3.3	2.2
		CI (+/-)	0.21	0.19	0.26	0.18	0.25	0.21	0.14	0.14
	0	Rating	3.8	2.9	2.3	1.5	2.8	2.2	2.9	2.2
		CI (+/-)	0.16	0.20	0.16	0.13	0.19	0.20	0.12	0.12
	5	Rating	4.3	3.4	3.3	2.6	3.4	3.0	3.7	3.0
		CI (+/-)	0.10	0.11	0.14	0.11	0.12	0.12	0.08	0.07
	10	Rating	4.1	3.8	3.3	3.3	3.4	3.3	3.6	3.5
		CI (+/-)	0.12	0.12	0.15	0.14	0.14	0.13	0.10	0.10

Table 5.6-3: Mean Ratings by D/U: Sky to Sky

			Delphi		GE		SONY		TOTAL	
	D/U		OFF	ON	OFF	ON	OFF	ON	OFF	ON
OFF Axis	-10	Rating	1.9	1.1	1.4	1.0	1.3	1.0	1.5	1.0
		CI (+/-)	0.19	0.09	0.12	0.00	0.11	0.05	0.09	0.03
	0	Rating	3.4	2.8	2.5	2.0	2.7	2.7	2.8	2.5
		CI (+/-)	0.20	0.16	0.23	0.16	0.19	0.18	0.13	0.11
	5	Rating	3.8	3.9	3.3	3.3	3.2	2.9	3.4	3.4
		CI (+/-)	0.18	0.16	0.18	0.16	0.19	0.18	0.11	0.11
	10	Rating	4.3	3.9	2.8	2.6	3.5	3.4	3.5	3.3
		CI (+/-)	0.13	0.18	0.26	0.24	0.16	0.19	0.13	0.13

Appendix A – Test Participant Questions

Participants will be asked a series of questions prior to the start of the test session. These questions are designed to elicit behavioral/preference information that may be incorporated in the final data analysis. The following questions will be posed to each test participant:

*Have you listened to **FM** radio within the last:*

- ❑ *Year*
- ❑ *Month*
- ❑ *Week*
- ❑ *Day*

*IF you listen to **FM** radio on a regular basis (even if only once in a while), Please answer the following questions:*

*How many hours per day do you listen to **FM** radio?*

- ❑ *Less than 30 minutes per day*
- ❑ *30 to 59 minutes per day*
- ❑ *1 to 2 hours per day*
- ❑ *More than 2 hours per day*

*Where do you spend most of your time listening to **FM** radio?*

- □ *In the car (on a car radio)*
- □ *At home (on a stereo)*
- ⧈ *At home (on a boombox)*
- ⧈ *At home (on a walkman)*
- ⧈ *Outdoors (on a boombox)*
- ⧈ *Outdoors (on a walkman)*
- ⧈ *At work (on a stereo)*
- ⧈ *At work (on a boom box)*
- ⧈ *At work (on a walkman)*
- ⧈ *In public places (i.e., gyms, malls, etc.)*
- ⧈ *Other* _____

*Have you listened to **AM** radio within the last:*

- ❑ *Year*
- ❑ *Month*
- ❑ *Week*

- ❑ *Day*

*IF you listen to **AM** radio on a regular basis (even if only once in a while), Please answer the following questions:*

*How many hours per day do you listen to **AM** radio?*

- ❑ *Less than 30 minutes per day*
- ❑ *30 to 59 minutes per day*
- ❑ *1 to 2 hours per day*
- ❑ *More than 2 hours per day*

*Where do you spend most of your time listening to **AM** radio?*

- □ *In the car (on a car radio)*
- □ *At home (on a stereo)*
- ❑ *At home (on a boombox)*
- ❑ *At home (on a walkman)*
- ❑ *Outdoors (on a boombox)*
- ❑ *Outdoors (on a walkman)*
- ❑ *At work (on a stereo)*
- ❑ *At work (on a boom box)*
- ❑ *At work (on a walkman)*
- ❑ *In public places (i.e., gyms, malls, etc.)*
- ❑ *Other* _____

*For both **AM** and **FM**, which types of radio shows do you listen to? (Check all that apply).*

- ❑ *Sports*
- ❑ *News*
- ❑ *Music*
- ❑ *Talk Shows*
- ❑ *NPR*
- ❑ *Religious*
- ❑ *Other* _____

Which types of music do you listen to? (Check all that apply!)

- ❑ *Alternative*
- ❑ *Classical*
- ❑ *Country*
- ❑ *Jazz/Blues*
- ❑ *New Age*
- ❑ *Pop/Rock*
- ❑ *Gospel*
- ❑ *Oldies*
- ❑ *Ethnic*

- ❑ *Rap*
- ❑ *R&B*
- ❑ *Other*
- ❑ *Classic Rock*

*What stations do you listen to during the day? (list the name or the "call numbers" and the locations they come from, if known)*_____

_____

*What stations do you listen to at night? (list the name or the "call numbers" and the locations they come from, if known)*_____

_____

Do you have favorite radio station(s)? List: _____

What are your favorite radio shows? List: _____

How often do you listen to them? (Check only one)

- ❑ *Daily*
- ❑ *Semi-Weekly*
- ❑ *Weekly*
- ❑ *Semi-Monthly*
- ❑ *Monthly*

Do you ever listen to stations that are outside of the DC/Baltimore area during the day?

- ❑ *Yes*
- ❑ *No*

*If yes, which ones?*_____

_____

Do you listen to far-away stations at night that you may not be able to hear during the day?

- ❑ *Yes*
- ❑ *No*

If "yes", name or list the "call numbers" for those stations (and cities they come from, if known:

_____

_____

*What is your biggest complaint about FM radio?*_____

*What is your biggest complaint about AM radio?*_____

*Do you work in the audio industry?*_____

*Do you work in the radio industry?*_____

*What kind of car do you own?*_____

Do you know what kind of radio is in your car? If yes, what is the brand? _____

_____

Do you have standard or special speakers in your car? _____

Appendix B – Experimenter Script

Welcome to our session! Today you will be participating in an audio test which should last approximately 2 hours. For this test you will hear approximately 250 short AM radio transmissions. Please listen to the clip from start to finish. Please listen only once. At the end of the clip, you will be asked to judge the sample on a 5-point scale. The clips you are going to hear are taken from news, sports and talk shows and from commercials. All of them are taken from real AM radio transmissions from different radio stations around the country. Once you start a session, you should continue until the program tells you to take your break, **but** you are also encouraged to take the test at your own pace. This may mean stopping between samples if you feel you need to "clear your head" for a few seconds.

For each sample, we ask you to keep the following scale in mind *(Experimenter – give participants a copy of the scale now)*:

Rating that you will see on the screen	**Description of Rating (as provided to test subjects)**
Unimpaired (Keep On)	This sample sounded good. I would listen to this audio under all circumstances.
Somewhat Impaired (Keep On)	This sample sounded good, but I heard some background impairments and noise. Still, I would listen to this audio a majority of the time.
Noticeably Impaired (Keep on if Motivated)	This sample was intelligible, but the background chatter and noise was noticeable and significant. I would continue to listen to this audio a majority of the time only if I was extremely interested in the program.
Severely impaired (Keep on only sometimes if extremely motivated)	This sample was mostly intelligible but the background chatter and noise was very annoying. I would continue to listen some of the time only if I was extremely interested in the program.
Failed (Turn off)	This sample is unintelligible. I would not listen to this audio under any circumstance.

You will have this scale with you on paper at all times, even though the screen will only display what is written in the left-hand column. Let's review the scale together. Notice that in a few categories (the top 2) you will be indicating that you would listen to the audio either under all circumstances or a majority of the time.

In other categories (the bottom 3) you will be indicating that you would listen to the audio under special circumstances – when you are motivated to listen. For example: suppose you were in your car listening to your favorite news show or sports broadcast. The program is one that you really are interested in and have been looking forward to hearing. It is unique – you can't get the same program from another channel (example: a college basketball game or a religious show that you know and like). You would use these categories to describe whether you would continue to listen to this special broadcast or whether you would try to find another station to listen to.

In all cases, we want to remind you that we are not asking you to judge the program material, or what's being talked about. We know that you will have various feelings about the sports and sports announcers, talk shows or commercials that you will hear. For this test, we are asking you to try to keep focused on only two things: (a) the quality of the transmission you are listening to and (b) the condition under which you are listening.

Appendix C – Ground to Sky Individual Sound Sample Ratings

				Delphi		GE		SONY		TOTAL		AVERAGE D/U LEVEL	
	D/U Level	Cut Number		OFF	ON	OFF	ON	OFF	ON	OFF	ON	OFF	ON
Off Axis	-5	1	Rating	3.70	2.84	3.47	3.14	3.58	3.07	3.58	3.02	-3.570	-4.140
			CI (+/-)	0.22	0.19	0.21	0.22	0.20	0.18	0.21	0.20		
	0	1	Rating	3.63	2.65	3.51	3.02	3.56	2.91	3.57	2.86	0.273	0.260
			CI (+/-)	0.21	0.20	0.21	0.23	0.24	0.21	0.22	0.22		
On Axis	-10	1	Rating	2.91	2.30	2.93	1.77	1.05	1.09	2.29	1.72	-11.942	-5.897
			CI (+/-)	0.26	0.15	0.25	0.17	0.06	0.09	0.34	0.20		
		4	Rating	1.60	1.58	1.70	1.33	1.02	1.00	1.44	1.30	-11.238	-7.180
			CI (+/-)	0.20	0.17	0.20	0.16	0.05	0.00	0.19	0.15		
	-5	1	Rating	3.74	2.12	3.19	2.14	2.91	2.44	3.28	2.23	-6.370	-4.973
			CI (+/-)	0.20	0.16	0.26	0.19	0.21	0.21	0.25	0.19		
		2	Rating	4.02	2.26	3.05	1.72	3.02	2.05	3.36	2.01	-5.605	-5.611
			CI (+/-)	0.19	0.20	0.26	0.21	0.22	0.17	0.26	0.20		
		4	Rating	2.98	2.60	2.60	2.72	1.14	1.60	2.24	2.31	-4.021	-2.690
			CI (+/-)	0.23	0.21	0.22	0.20	0.10	0.17	0.30	0.24		
		5	Rating	2.60	1.16	1.60	1.00			2.10	1.08	-6.290	-8.047
			CI (+/-)	0.24	0.13	0.20	0.00			0.26	0.09		
	0	1	Rating	3.67	3.00	2.65	2.56	3.77	3.16	3.36	2.91	0.428	1.296
			CI (+/-)	0.18	0.19	0.24	0.25	0.24	0.17	0.27	0.22		
		2	Rating	4.05	2.95	3.86	2.60	2.67	1.58	3.53	2.38	3.138	1.356
			CI (+/-)	0.16	0.20	0.26	0.19	0.22	0.16	0.28	0.25		
		3	Rating	2.72	3.12	1.95	2.44	2.23	2.63	2.30	2.73	-1.750	1.322
			CI (+/-)	0.25	0.16	0.17	0.22	0.17	0.24	0.22	0.23		
	10	2	Rating	3.44	3.14	2.19	2.37	2.12	2.44	2.58	2.65	6.840	12.080
			CI (+/-)	0.21	0.25	0.22	0.22	0.20	0.19	0.27	0.24		

Appendix D – Sky to Ground Individual Sound Sample Ratings

				Delphi		GE		SONY		TOTAL		AVERAGE D/U LEVEL	
	D/U Level	Cut Number		OFF	ON	OFF	ON	OFF	ON	OFF	ON	OFF	ON
Off Axis	-5	1	Rating	1.88	1.67	3.09	3.42	3.00	2.88	2.66	2.66	0.073	-5.420
			CI (+/-)	0.17	0.19	0.23	0.24	0.21	0.22	0.26	0.31		
	0	1	Rating	2.67	3.37	3.56	2.56	3.42	2.63	3.22	2.85	1.202	0.222
			CI (+/-)	0.23	0.26	0.26	0.23	0.28	0.23	0.28	0.26		
	5	2	Rating	3.84	2.88	3.58	3.33	3.72	3.49	3.71	3.23	3.790	2.528
			CI (+/-)	0.25	0.20	0.21	0.22	0.25	0.30	0.24	0.25		
	10	2	Rating	3.42	3.86	3.33	3.63	3.12	3.53	3.29	3.67	13.062	11.518
			CI (+/-)	0.20	0.21	0.22	0.24	0.26	0.25	0.23	0.24		
On Axis	-5	1	Rating	3.49	2.51	3.00	1.49	3.56	2.53	3.35	2.18	-3.669	-6.059
			CI (+/-)	0.21	0.19	0.26	0.18	0.25	0.21	0.25	0.24		
	0	2	Rating	3.70	2.26	2.16	1.28	2.35	1.47	2.74	1.67	3.620	1.113
			CI (+/-)	0.27	0.22	0.23	0.15	0.20	0.15	0.31	0.21		
		4	Rating	3.84	3.49	2.40	1.63	3.16	2.84	3.13	2.65	6.570	1.215
			CI (+/-)	0.18	0.20	0.21	0.19	0.27	0.22	0.28	0.31		
	5	1	Rating	3.93	3.16	3.09	2.47	3.26	2.98	3.43	2.87	6.160	4.040
			CI (+/-)	0.19	0.20	0.25	0.20	0.21	0.21	0.24	0.22		
		2	Rating	4.05	3.26	2.88	2.40	3.30	2.60	3.41	2.75	5.650	8.006
			CI (+/-)	0.21	0.17	0.30	0.21	0.23	0.21	0.29	0.22		
		3	Rating	4.66	3.57	3.67	2.85	3.62	3.24	3.98	3.22	1.630	3.300
			CI (+/-)	0.16	0.23	0.26	0.23	0.24	0.25	0.26	0.25		
	10	2	Rating	3.95	3.35	2.72	2.74	3.21	3.00	3.29	3.03	6.520	10.337
			CI (+/-)	0.20	0.19	0.25	0.22	0.26	0.21	0.28	0.22		
		3	Rating	3.98	4.05	3.42	3.37	3.44	3.40	3.61	3.60	7.870	11.680
			CI (+/-)	0.20	0.20	0.24	0.21	0.25	0.22	0.24	0.23		
		5	Rating	4.28	4.12	3.77	3.72	3.60	3.58	3.88	3.81	14.310	14.530
			CI (+/-)	0.19	0.16	0.19	0.20	0.20	0.22	0.21	0.20		

Appendix E – Sky to Sky Individual Sound Sample Ratings

			Delphi		GE		SONY		TOTAL		AVERAGE D/U LEVEL	
D/U Level	Cut Number		OFF	ON	OFF	ON	OFF	ON	OFF	ON	OFF	ON
-10	2	Rating	2.60	1.12	1.65	1.00	1.63	1.00	1.96	1.04	-9.970	-14.150
		CI (+/-)	0.20	0.19	0.17	0.00	0.18	0.00	0.23	0.11		
	3	Rating	1.19	1.00	1.09	1.00	1.00	1.05	1.09	1.02	-19.416	-16.094
		CI (+/-)	0.13	0.00	0.11	0.00	0.00	0.09	0.10	0.05		
0	2	Rating	2.95	2.63	1.63	1.58	2.12	2.16	2.23	2.12	-2.405	-1.622
		CI (+/-)	0.26	0.22	0.17	0.20	0.21	0.21	0.27	0.24		
	3	Rating	3.84	2.95	3.35	2.51	3.19	3.16	3.46	2.88	-2.654	-1.184
		CI (+/-)	0.23	0.23	0.22	0.18	0.23	0.21	0.24	0.22		
5	2	Rating	3.86	4.19	3.35	3.42	3.23	2.93	3.48	3.51	4.900	8.951
		CI (+/-)	0.25	0.22	0.25	0.22	0.27	0.25	0.27	0.28		
	3	Rating	3.81	3.67	3.21	3.12	3.12	2.81	3.38	3.20	6.805	5.070
		CI (+/-)	0.28	0.21	0.26	0.23	0.26	0.25	0.28	0.25		
10	3	Rating	4.49	4.23	3.81	3.49	3.56	3.65	3.95	3.79	12.620	11.695
		CI (+/-)	0.18	0.23	0.22	0.28	0.25	0.29	0.24	0.28		
	4	Rating	4.14	3.60	1.79	1.74	3.44	3.12	3.12	2.82	8.696	6.819
		CI (+/-)	0.18	0.24	0.21	0.16	0.21	0.22	0.35	0.31		