Report On Perceptual Tests of Coders at Low- and Very Low-Bit Rates

Prepared for National Public Radio & IAAIS
In Cooperation with
iBiquity Digital Corporation

by
Ellyn G. Sheffield, PhD
Sheffield Audio Consulting
Princeton, New Jersey 08540
(609) 737-9451
egsheffield@comcast.net

1 Introduction

The primary goal of this study was to determine whether the perceived audio quality of future digital coders running at low bit rates (under 25kbps) was comparable to or better than FM SCA transmissions currently available to reading services consumers. Approximately 1 million visually impaired or blind consumers rely on the programming provided by Reading Services. This programming is varied, including speech, music, and voiceovers. The quality of existing analog SCA service varies greatly, depending on the reception areas in which consumers live. Thus, in order to comprehensively test digital against existing analog service, both unimpaired and impaired analog transmissions needed to be included as benchmarks.

In two tests, participants were asked to judge the quality of several coders on a number of dimensions:

- The overall quality (on a five-point scale; Bad to Excellent)
- Discernable impairments (on a six-point scale; None heard to Extremely annoying)
- Satisfaction with the transmission (Yes/No)

During the first test, participants listened to short, 15-second samples and during the second test participants listened to longer samples (approximately 35-75 seconds). By comparing the same participants' ratings between short and long samples, we were able to identify whether their opinion of sample quality changed after longer exposure times, providing a quantifiable indicator of listener fatigue.

Audio source material was created by National Public Radio (NPR) and distributed to individual U.S., Canadian and European companies as part of an RFI, attached as Appendix 1. Vendors were given free reign to decide on coding parameters (i.e., bit-rate; processing), as long as they documented and returned their submission to NPR. All material coded between 0-11.8 kbps was included in a "very low bit rate" (VLBR) category, and material coded between 12-25 kbps was included in a "low bit rate"(LBR) category. These data-rate groups were chosen based on the payload available in a single as well as two extended hybrid partitions available for use in iBiquity's IBOC DAB radio system.

Broadcasting services provided on FM subcarriers provide the basis for most reading services for the sight impaired.[1], therefore the experimental methodology gave special consideration to this population. The participant population included equal numbers of visually impaired persons (VIPs) and non-visually impaired persons (NVIPs). All participants were tested using methodologies created especially for VIPs (a detailed description of the methodology is given in Section 2).

The study was conducted in two phases during the months of July and August. The first phase narrowed the field of coders in order to limit the number of test conditions ultimately presented to general public listeners. This phase was conducted with a small number of experts from the radio industry, including audio engineers, CEA, NAB and NPR personnel. The second phase

[1] According to data provided by the International Association of Audio Information Services (IAAIS) well over 100 radio reading services are in daily operation in the United States.

was designed to obtain ratings from the general public (both VIPs and NVIPs) on three VLBR and three LBR coders. The details of both test phases are described in the remaining sections of this report.

2 Test Methodology

2.1 Sound Samples – Source Material

Male and female speech, male and female voice-overs, and music (rock, jazz and classical) were included in this study. Audio selections were chosen from a variety of sources, including speech passages recorded by readers from Reading Services, music CD's and voiceovers from Sun Sounds of Arizona program material. Reading passages were selected from articles found in newspapers, including the Princeton Packet, Princeton, New Jersey; Trenton Times, Trenton, New Jersey; Washington Post, Washington DC; and the New York Times, New York NY. In order to ensure that listeners would not be influenced by the programmatic content when listening for audio quality, special care was taken to include passages that were interesting but not controversial. Therefore, testable material included travel, food, housing, and lifestyle articles, while material referencing political, violent, biased, or otherwise potentially controversial subjects was excluded. Passages were sent from NPR to the International Association of Audio Information Services (IAAIS) and were recorded by volunteer readers at several locations. Three female and three male readers were chosen to record the passages. These voices were chosen based on their pitch, timbre, accent, comprehensibility, and overall quality. The final test material exhibited some minor quality differences as a result of recordings being produced using different readers and studios. These differences accurately simulated the varying fidelity of typical, real-world Reading Services programs. Additionally, voiceovers contributed by Sun Sounds of Arizona were indicative of normal Reading Services program material. Therefore we believe the resulting speech and voiceover material was appropriate for test purposes and indicative of the type of source material coders may be required to handle in the future.

2.2 Preparation of SCA Audio Samples

Audio samples were prepared on the FM test bed shown in Figure 2.3. The system produced a standard FM-band signal with stereo and 67 kHz SCA subchannels in compliance with the FCC rules, 47 CFR 73.322 and 47 CFR 73.319.

The test bed passed audio samples from an audio CD through a transmission/receiving chain. The resulting SCA subchannel audio was recorded on audio CD, for later transfer to playback equipment used by the listeners.

Audio samples were processed in a manner typical of FM subchannel operations, such as radio reading services. A Moseley TFL-280 using 5 kHz low-pass filtering and 150 µS pre-emphasis provided peak limiting and compression. The RCA 67 kHz subcarrier generator was to a maximum of 6 kHz peak deviation. Its output was connected through an Omnia-6EX-HD, which combines the 67 kHz subcarrier with the internal stereo generator.

The Omnia also performed processing of Main Channel audio for the stereo generator, to simulate typical broadcast program conditions, which created Main-to-Subcarrier crosstalk for the “McMartin Light” and “McMartin Medium” impaired SCA audio samples. Audio material on the playback CD was extracted from the voiceover samples used in the perceptual testing, as this voice-and-music combination provided generic aural source material. An audio sample of approximately 30 seconds duration was repeated automatically by the CD player to provide a consistent source of main channel modulation.

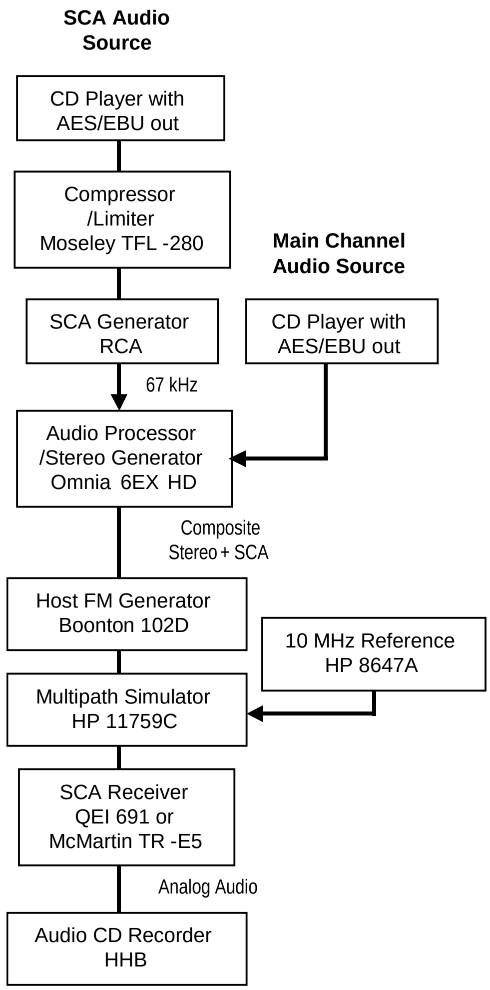


Figure 2.3 –Equipment configuration for preparation of audio samples

2.3 Test environment

Testing was conducted in two, production-quality sound studios at National Public Radio, Washington, DC. The studios were approximately 9 x 9 feet, and acoustically identical.

Audio samples were presented to listeners monaurally over one Genelec™ near field monitor. For Phase 1, listeners were tested individually. For Phase 2, 3 listeners were placed approximately 7 feet from the monitor, as recommended by ITU recommendation BS-1116.

Participants were seated next to each other, separated by opaque, dark curtains. While these curtains did not interfere with the audio, they blocked NVIPs from viewing each other's responses.

3 Phase 1 - Narrowing the field of coders

In order to bring the total number of vendor submissions to three very low bit rate coders (VLBR) and three low bit rate coders (LBR), a panel of experts working in the audio industry evaluated all submissions. See Appendix 2 for experts and their affiliations.

3.1 Methodology

Using an adapted version of ITU-R recommended MUSHRA methodology, audio experts ranked coders in each of five genres: (a) male and female speech; (b) male and female voiceover; (c) rock; (d) jazz; (e) classical music. Coders were not named or identified in any discernable way during this process.

For each sample, participants were presented with a list of coded wave files through Windows Media Player. They were instructed to play the list as many times as necessary to rate each coder. Their job was to rate each coder independently, while keeping in mind that they were also "rank-ordering" all coders. Thus, they were instructed that each coder should be assigned a unique score, even if the coders sounded very similar. For fine discrimination, participants were allowed to rate each sample on a 50-point scale, anchored at increments of 10. This followed the ACR-MOS scale with the exception that participants were able to rate the samples as "failed" (0). Therefore, participants were able to assign a number between 0 – 50, 0 = "failure", 10 = "bad", 20 = "poor", 30 = "fair", 40 = "good", and 50 = "excellent". Appendix 3 is the Experimenter script used to explain the task to experts and shows the answer sheet they used to register their responses.

3.2 Results

VLBR and LBR ratings were calculated separately. Results from individual rankings were tabulated, and overall scores were calculated as follows:

- Except for "speech" all 8 participants' ratings were added to obtain a score for each category.
- In "male speech" and "female speech", participants' ratings were doubled to obtain the score.
- Scores were totaled and three coders were chosen in each category based on the highest scores.

Table 3.2.1 and Table 3.2.2 show the ratings for the LBR and VLBR categories respectively. Letters identify all coder vendors who were not ultimately used in Phase 2 consumer testing.

Table 3.2.1 – Low Bit Rate Total Scores

	A	B	C	iBiquity	E	F	Orban	H	VoiceAge
Classical	113	139	268	258	263	267	283	96	250
Jazz	135	215	286	312	249	222	299	129	256
Female Speech	352	436	532	530	542	316	516	406	542
Male Speech	386	354	472	524	528	354	474	384	552
Rock	103	211	287	294	209	203	280	102	272
Voice Over	158	260	286	281	247	209	292	212	294
TOTAL	1247	1615	2131	**2199**	2038	1571	**2144**	1329	**2166**

Table 3.2.2 - Very Low Bit Rate Total Scores

Data	A	B	C	D	iBiquity	Anon-ymous	G	H	I	VoiceAge
Classical	50	61	31	238	232	282	207	221	59	266
Jazz	64	66	45	245	223	145	155	240	70	284
Female Speech	356	406	484	188	360	528	348	182	320	570
Male Speech	246	334	470	144	284	498	312	158	286	550
Rock	66	81	101	268	248	102	131	211	64	229
Voice Over	159	193	194	192	221	221	140	168	149	265
TOTAL	941	1141	1325	1275	**1568**	**1776**	1293	1180	948	**2164**

Based on expert rankings, iBiquity's HDC, Anonymous and Voice Age were included in Phase 2 testing for very low bit rates, and iBiquity's HDC, Orban and Voice Age coders were included for low bit rates.

4 Phase 2 – Consumer Test

4.1 Participants

Listeners were recruited from several sources. VIPs were recruited from across the country by IAAIS, the American Council of the Blind (ACB), and the National Federation of the Blind (NFB). Sighted listeners were either associated with IAAIS, or recruited from contacts with NPR staff, flyers posted in the downtown Washington area and outlying suburbs, and on-line postings. Sixty-one listeners (35 VIPs and 26 NVIPs) participated, distributed between 18 and 75 years of age. Subjective data from 49 listeners was collected (25 VIPs and 24 NVIPs), where qualification was based on performance on the initial screening test. Seven VIPs and 1 NVIP were excluded from final results because they failed the screening test and/or did not complete the test (see Section 4.3.1 for details of the screening test). Two VIPs and one NVIP were excluded due to experimenter error. Table 4.1 shows the demographic breakdown of listeners.

Table 4.1 - Demographic breakdown of listeners

	VIP		NVIP	
	Male	Female	Male	Female
18-30	2	3	4	4
31-45	4	3	3	2
46-60	3	4	3	3
61+	3	3	2	3

4.2 Sound Samples

Speech, voiceover and music samples were included in the test. For the short sample test, two samples were included from each speech category and voiceover segment. All samples were 10-15 seconds long, and were chosen based on their programmatic content (neutral being best), overall quality and appeal. Music samples included representative samples from classical, rock and jazz genres. Source recordings were included to ensure that participants heard high-quality audio samples from time to time during the test.

Long samples used in "fatigue" testing were approximately 35 to 75s in length. The short 10-15 second segments were imbedded in these longer samples. By using expanded material for the long test samples, it was possible to directly compare participants' ratings of short and long samples to see if the scores were fundamentally similar or different. See Appendix 5 for samples of the scripts used in both short and long segments.

4.3 Design and procedures

Listeners participated in three listening tests: a short screening test and two main tests, each of which lasted approximately about 2½ hours. Three participants, considered a test panel, were tested at one time. The order of testing was constant; that is all participants were screened, participated in the short-sample test, took a 1-2 hour break and then participated in the long-sample test. The order of audio sample presentation was randomized within test panels. Therefore, each test panel received a different sample presentation order for each test.

4.3.1 Screening

Screening was conducted to ensure that listeners were reliably able to distinguish between significantly different audio qualities. There were 7 screening trials. For each trial, participants were asked to listen to 3 samples, 2 of which were the same and the 3rd different (for example, 2 female speech source samples and the same female speech sample processed through an AM receiver; 2 rock source samples from a CD and the third sample coded at HDC 24 kbps). The listener's task was to decide which of two "test" samples ("A" or "B") was different from the reference sample. In each trial, the first sample they heard was always the "reference" sample. They then listened to the "A" and "B" samples and judged which of the samples was different from the reference. Listeners were free to ask the experimenter to replay any or all of the three samples until they were ready to respond and proceed to the next trial. Listeners were asked to hold up 1 finger if sample "A" was different from the reference, and hold up 2 fingers if sample "B" was different from the reference. In order to pass the screening test, participants had to answer 5 of the 7 triads correctly. Listeners were provided no feedback on the "correctness" of their responses during the screening test nor were they informed of their specific performance after they were finished.

4.3.2 Main tests: Short Sample and Long Sample

The short-sample test included 144 short samples, and the long-sample test included 70 long samples. See.Table 4.3.2.1 and Table 4.3.2.2 for details of the design. To ensure that listening to the same samples did not overly fatigue participants, different sub-segments of audio were recorded through VLBR and LBR coders for Classical, Jazz, Rock and Voiceover (for example, if the LBR Classical sample was taken from the first 15 seconds of the Ibert Oboe concerto, the VLBR Classical sample was taken from a later 15 second segment of the same movement). In order to directly compare digital to analog audio, both of the VLBR and LBR sub-segments were recorded through the QEI, McMartin light and McMartin medium processors. Therefore, the number of analog recordings was doubled in the classical, jazz, rock and voiceover genres. However, the same speech samples were used for VLBR and LBR digital recordings; thus the same number of digital and analog recordings was made.

Table 4.3.2.1 - Test 1: Short Sample Design

Coder/Bit Rate	Classical	Jazz	Rock	Speech	Voiceover	Total
iBiquity LBR	1	2	1	6	2	12
Orban LBR	1	2	1	6	2	12
VoiceAge LBR	1	2	1	6	2	12
iBiquity VLBR	1	2	1	6	2	12
Anonymous VLBR	1	2	1	6	2	12
VoiceAge VLBR	1	2	1	6	2	12
QEI Unimpaired	2	4	2	6	4	18
McMartin Light	2	4	2	6	4	18
McMartin Medium	2	4	2	6	4	18
Source	2	4	2	6	4	18
Total						**144**

As Table 4.3.2.2 shows, 70 recordings were included in the long sample test. Due to the length of the test samples and the fatiguing nature of the task, we included only 4 of 5 genres for low bit rate testing and 4 of 5 genres for very low bit rate testing.

Table 4.3.2.2 - Test 2 - Long Sample Design

Coder/Bit Rate	Classical	Jazz	Rock	Speech	Voiceover	Total
iBiquity LBR	0	1	1	2	1	5
Orban LBR	0	1	1	2	1	5
VoiceAge LBR	0	1	1	2	1	5
iBiquity VLBR	1	1	0	2	1	5
Anonymous VLBR	1	1	0	2	1	5
VoiceAge VLBR	1	1	0	2	1	5
QEI Unimpaired	1	2	1	4	2	10
McMartin Light	1	2	1	4	2	10
McMartin Medium	1	2	1	4	2	10
Source	1	2	1	4	2	10
Total						**70**

Regardless of whether participants were VIPs or NVIPs, they were given identical instructions. Participants listened to samples, one-by-one, and answered several questions about each sample. For each question, after listening carefully to the sound sample, participants were told to raise their fingers in front of their chest until the experimenter recorded their answers. Participants rated the following questions:

- What was the “overall quality” of the sound sample? Participants used the 5-point ACR-MOS scale to rate quality.
- If audio impairments were heard, how annoying were they? Impairments were defined as odd sounds, annoying artifacts, things that just sounded strange. Participants used a 6-point scale, ranging from “no impairments” to “heard impairments, and they were extremely annoying”.
- How satisfied were you with the quality? That is, would you keep the station on or turn it off because of the quality? Participants were asked to think of a normal listening situation before they answered (i.e., they were interested in the show, but not “desperate” to hear the program). Participants used a two-point scale, “keep the radio on”/ “turn the radio off”.

Data collection software was created at NPR specifically for this use. See Appendix 4 for the Experimenter script used for all participants.

4.4 Preliminary Results

Preliminary analyses were conducted to examine whether participants rated audio quality of samples differently based on their age or gender. A 2 (gender) x 4(age) ANOVA yielded a main effect of age, a main effect of gender, and an interaction. Newman-Keuls Multiple-Comparison tests indicated that, as with past audio testing, older participants rated samples less critically than younger participants. Additionally, males rated samples less critically than females. However, this effect was seen primarily with males over the age of 60. That is, 60+ males rated samples significantly higher than did 60+ females (3.91 and 3.42 respectively).

Another analysis was conducted to examine whether visually impaired listeners rated samples differently from sighted listeners. This analysis yielded no main effect of “sight” classification, suggesting that visually impaired and sighted listeners rated samples statistically similarly. Because there was no effect, all respondents were combined for future analyses. See Appendix 6 for scores divided by sight classification.

4.5 Results – Quality Scores

Although VLBR and LBR samples were included in the same short-sample test, resulting data from were analyzed separately. Figure shows the results of quality scores for low bit rates, divided by “music”, “speech”, and “voiceover”. Music is a combination of classical, jazz and rock selections. At low bit rates, all digital coders were rated superior to SCA transmissions. Orban and iBiquity were strongest in music and VoiceAge was strongest in speech with participants rating it equivalently to source material. Coders performed equally well in the Voiceover category.

As seen in Figure 4.5.2 at very low bit rates, only VoiceAge performed either at or significantly better than QEI. iBiquity and VoiceAge were rated equivalently to QEI in the music category, but only VoiceAge out performed QEI in speech. Details for quality ratings can be found in Table 4.5.1 through Table 4.5.7.

Figure 4.5.1 – Quality Ratings at Low Bit-Rates for Music Speech and Voiceover.

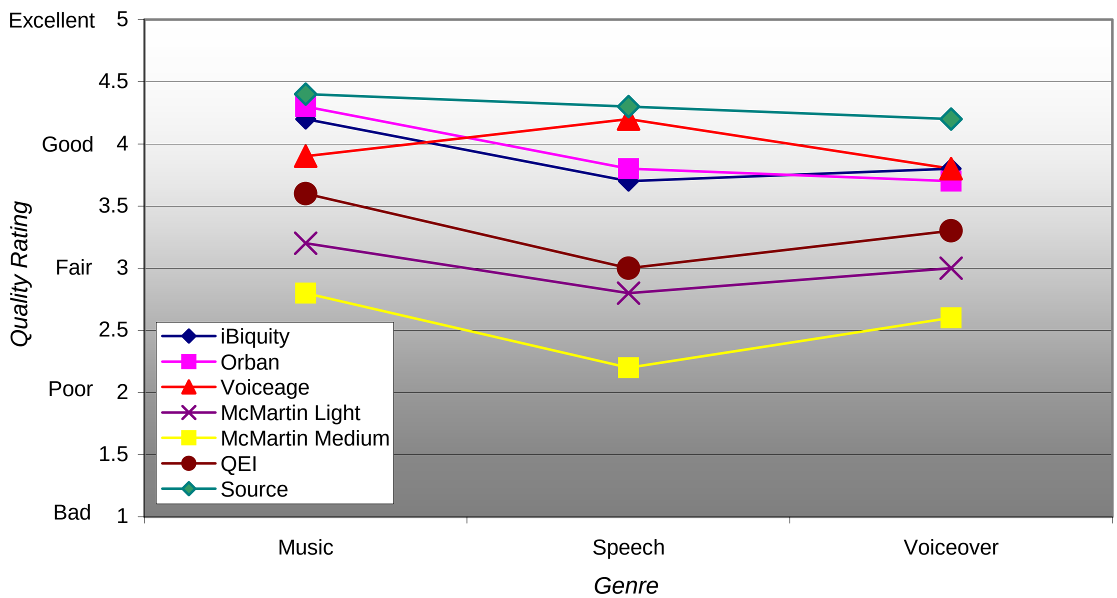


Figure 4.5.2 - Quality Ratings at Very Low Bit-Rates for Music, Speech and Voiceover.

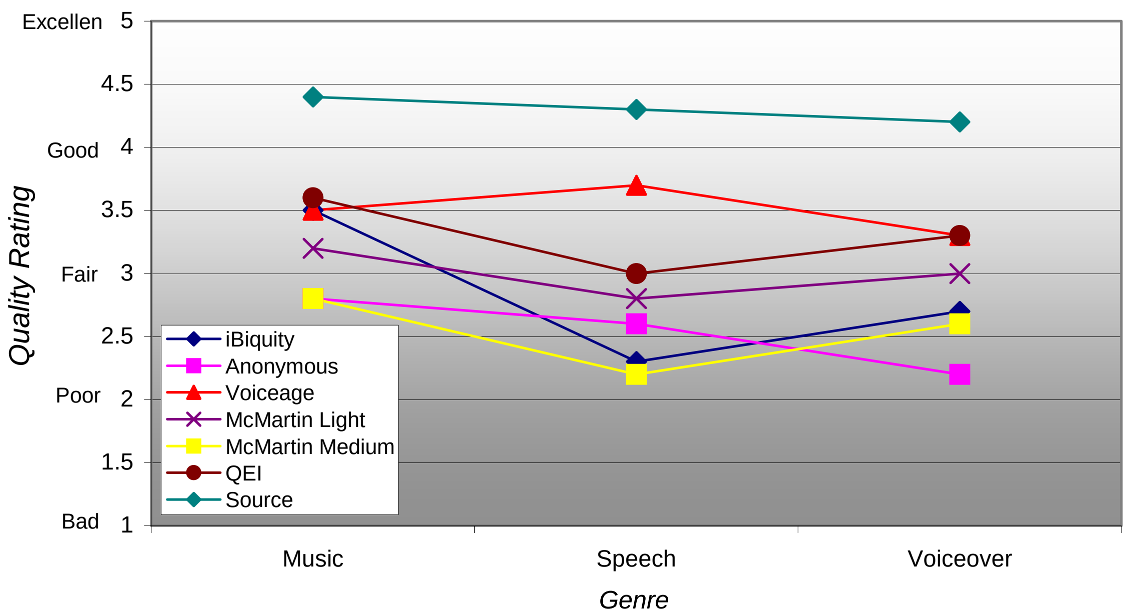


Table 4.5.1 shows the breakdown of ACR Quality MOS scores for all low bit rates. The results are listed by genres. A one-way analysis of variance (ANOVA) was conducted for each genre to see if the scores among coders were significantly different from each other. These results yielded significant differences, which are highlighted on Table 4.5.2 and Table 4.5.3. When compared to the QEI, iBiquity and Orban's coders were rated significantly higher in all genres except Voiceover and VoiceAge was rated significantly higher than QEI in speech. When compared to the McMartin Light, all digital coders were rated higher in all categories.

Table 4.5.1 - Overall Quality: ALL LOW BIT RATE SCORES

LBR	**Classical**	**Jazz**	**Rock**	**Speech**	**VoiceOver**
iBiquity	4.2	4.2	3.9	3.7	3.7
Orban	4.4	4.2	4.1	3.7	3.6
VoiceAge	4.0	4.0	3.4	4.1	3.7
QEI	3.6	3.7	3.4	3.1	3.4
McMartin Light	3.3	3.4	2.8	2.9	3.0
McMartin Med	2.7	2.9	2.7	2.2	2.4
Source	4.5	4.4	4.2	4.4	4.1

Table 4.5.2 - Low Bit Rate Overall Quality (Coders compared to QEI Unimpaired)

LBR	**Classical**	**Jazz**	**Rock**	**Speech**	**VoiceOver**
iBiquity	4.2*	4.2*	3.9*	3.7*	3.7
Orban	4.4*	4.2*	4.1*	3.7*	3.6
VoiceAge	4.0	4.0	3.4	4.1*	3.7
QEI	**3.6**	**3.7**	**3.4**	**3.1**	**3.4**
Source	4.5	4.4	4.2	4.4	4.1

*Significantly better than QEI

Table 4.5.3 - Low Bit Rate Overall Quality (Coders compared to McMartin Light Impairment)

LBR	**Classical**	**Jazz**	**Rock**	**Speech**	**VoiceOver**
iBiquity	4.2*	4.2*	3.9*	3.7*	3.7*
Orban	4.4*	4.2*	4.1*	3.7*	3.6*
VoiceAge	4.0*	4.0*	3.4*	4.1*	3.7*
McMartin Light	**3.3**	**3.4**	**2.8**	**2.9**	**3.0**
Source	4.5*	4.4*	4.2*	4.4*	4.1*

*Significantly better than McMartin Light

Table 4.5.4 shows the breakdown of ACR Quality MOS scores for all very low bit rates. Again, the results are listed by genres. A one-way analysis of variance (ANOVA) was conducted for each genre to see if the scores among very low bit rate coders were significantly different from each other. These results yielded significant differences, which are highlighted on Table 4.5.4, Table 4.5.5 and Table 4.5.6. When compared to the QEI and McMartin Light, only the VoiceAge coder was rated significantly higher in speech and classical. VoiceAge was no different than QEI and McMartin Light in all other categories. In contrast, iBiquity and Anonymous were rated either at or lower than QEI, and iBiquity was rated lower than McMartin Light in speech. When compared to McMartin Medium, VoiceAge and Anonymous were rated significantly higher in speech and classical, and iBiquity was rated significantly higher in classical and jazz.

Table 4.5.4 - Overall Quality: All Very Low Bit Rate Scores

LBR	Classical	Jazz	Rock	Speech	VoiceOver
iBiquity	3.9	3.6	3.1	2.1	2.8
Anonymous	4.0	2.9	2.5	2.7	2.4
VoiceAge	4.2	3.3	3.0	3.8	3.3
QEI	**3.6**	**3.7**	**3.4**	**3.1**	**3.4**
McMartin Light	3.3	3.4	2.8	2.9	3.0
McMartin Med	2.7	2.9	2.7	2.2	2.4
Source	4.5	4.4	4.2	4.4	4.1

Table 4.5.5 - Very Low Bit Rate Overall Quality (Coders compared to QEI Unimpaired)

LBR	Classical	Jazz	Rock	Speech	VoiceOver
iBiquity	3.9	3.6	3.1	2.1**	2.8**
Anonymous	4.0	2.9**	2.5**	2.7**	2.4**
VoiceAge	4.2*	3.3	3.0	3.8*	3.3
QEI	**3.6**	**3.7**	**3.4**	**3.1**	**3.4**
Source	4.5*	4.4*	4.2*	4.4*	4.1*

* Significantly greater than QEI Unimpaired
** Significantly less than QEI Unimpaired

Table 4.5.6 - Very Low Bit Rate Overall Quality (Coders compared to McMartin Light)

VLBR	Classical	Jazz	Rock	Speech	VoiceOver
iBiquity	3.9*	3.6	3.1	2.1**	2.8
Anonymous	4.0*	2.9**	2.5	2.7	2.4**
VoiceAge	4.2*	3.3	3.0	3.8*	3.3
McMartin Light	**3.3**	**3.4**	**2.8**	**2.9**	**3.0**
Source	4.5*	4.4*	4.2*	4.4*	4.1*

* Significantly greater than McMartin Light
** Significantly less than McMartin Light

Table 4.5.7 - Very Low Bit Rate Overall Quality (Coders compared to McMartin Medium)

LBR	Classical	Jazz	Rock	Speech	VoiceOver
iBiquity	3.9*	3.6*	3.1	2.1	2.8
Anonymous	4.0*	2.9	2.5	2.7*	2.4
VoiceAge	4.2*	3.3	3.0	3.8*	3.3*
McMartin Med	**2.7**	**2.9**	**2.7**	**2.2**	**2.4**
Source	4.5*	4.4*	4.2*	4.4*	4.1*

* Significantly greater than McMartin Medium
** Significantly less than McMartin Medium

4.6 Short listening time vs. longer listening time

In order to explore whether listeners changed their quality ratings over time, direct comparisons were made between the mean score of short samples (test 1) and long samples (test 2). A 2 (sample length) x 7 (condition) ANOVA was conducted with sample length and coders as within-participant factors. Results indicate that participants changed their ratings very infrequently, with most changes occurring in the very low bit rate classification (see Table 4.6

and Table 4.7). Thus, in general it appears that listeners made decisions about quality fairly early in the listening experience (within the first 10-15 seconds) and did not change their opinions after they listened for extended periods.

Table 4.6 - Short samples vs. Long Samples for Low Bit Rate Coders

	JAZZ		ROCK		SPEECH		VOICEOVER	
Coder	**MOS Short**	**MOS Long**	**MOS Short**	**MOS Long**	**MOS Short**	**MOS Long**	**MOS Short**	**MOS Long**
Orban	4.2	4.4	4.1	4.1	3.7	3.6	3.6	3.7
IBiquity	4.2	4.4	3.9	4.0	3.7	3.7	3.7	3.8
VoiceAge	4.0	4.1	3.4	3.7	4.1	4.0	3.7	3.9
QEI	3.7	3.7	3.4	3.3	3.1	3.2	3.4	3.3
McMartin Light	3.4	3.2	2.8	2.9	2.9	2.6**	3.0	3.0
McMartin Medium	2.9	2.8	2.7	2.5	2.2	2.1	2.4	2.4
Source	4.4	4.6	4.2	4.1	4.4	4.3	4.1	4.1

** Short greater than Long

Table 4.7 - Short samples vs. Long Samples for Very Low Bit Rate Coders

	CLASSICAL		JAZZ		SPEECH		VOICEOVER	
Coder	**MOS Short**	**MOS Long**	**MOS Short**	**MOS Long**	**MOS Short**	**MOS Long**	**MOS Short**	**MOS Long**
Anonymous	4.0	3.8	2.9	2.3**	2.7	2.9	2.4	1.9**
IBiquity	3.9	4.1	3.6	3.7	2.1	2.0	2.8	2.8
VoiceAge	4.2	4.1	3.3	3.2	3.8	4.1*	3.3	3.3
QEI	3.6	3.7	3.7	3.7	3.1	3.2	3.4	3.3
McMartin Light	3.3	3.3	3.4	3.2	2.9	2.6**	3.0	3.0
McMartin Medium	2.7	2.6	2.9	2.8	2.2	2.1	2.4	2.4
Source	4.5	4.2	4.4	4.6	4.4	4.3	4.1	4.1

*Long greater than Short
** Short greater than Long

4.7 Results - Impairments

Table 4.7.1 and Table 4.7.2 show participants' ratings of the level of impairments they heard on the recordings. These scores indicate that at low bit rates participants heard significantly more impairment on the QEI than they did for all the digital coders in the speech and classical genres, and in jazz the Orban was rated significantly better than QEI. However, reported impairments were equal in the voiceover category. At very low bit rates, significantly fewer impairments were heard with VoiceAge in the speech and classical genres, but significantly more impairments were heard with iBiquity and Anonymous in speech. In voiceover, only VoiceAge was equal to QEI; listeners heard significantly more impairments on both iBiquity and Anonymous than they did on QEI. These results corroborate quality ratings, indicating that, at least for speech, at very low bit rates listeners were significantly more satisfied with VoiceAge than all other systems.

Rating Scale
0 = Heard no impairments
1 = Heard impairments, not annoying
2 = Heard impairments, slightly annoying
3 = Heard impairments, annoying
4 = Heard impairments, Very annoying
5 = Heard impairments, Extremely annoying

Table 4.7.1 - Impairments Reported at Low Bit Rates (compared to QEI)

	Classical	Jazz	Rock	Speech	VoiceOver
iBiquity	.9**	.9	1.2	1.4**	1.5
Orban	.4**	.7**	1.0**	1.5**	1.5
VoiceAge	.9**	1.2	1.9	.9**	1.4
QEI	1.4	1.2	1.5	2.1	1.7
McMartin Light	2.1*	1.7*	2.4*	2.3	2.3*
McMartin Med	3.0*	2.4*	2.5*	3.3*	3.0*
Source	.6**	.6**	.9**	.8**	1.2**

** Significantly less impairment heard than QEI
* Significantly more impairment heard than QEI

Table 4.7.2 - Impairments Reported at Very Low Bit Rates (compared to QEI)

	Classical	Jazz	Rock	Speech	VoiceOver
iBiquity	1.3	1.5	2.2*	3.0*	2.5*
VoiceAge	.4**	1.9*	2.4*	1.3**	2.0
Anonymous	1.3	2.6*	3.6*	2.4*	3.1*
QEI	1.4	1.2	1.5	2.1	1.7
McMartin Light	2.1*	1.7*	2.4*	2.3	2.3*
McMartin Med	3.0*	2.4*	2.5*	3.3*	3.0*
Source	.6**	.6**	.9**	.8**	1.2**

** Significantly less impairment heard than QEI
* Significantly more impairment heard than QEI

Table 4.7.3 and table 4.7.4 compare short and long listening tests. In low bit rate digital conditions, participants did not hear more impairment over time. In very low bit rate conditions participants claimed that they heard more impairments in 4 digital conditions (VoiceAge – Classical; Anonymous – Jazz and Voiceover, and iBiquity – Speech). In SCA conditions participants heard more impairments both in McMartin Light and McMartin Medium. Note that in the majority of these cases, the impairments had already reached the “annoying level” in short sample listening and merely seemed to become more noticeable over time.

Table 4.7.3 - Impairments Reported at Low Bit Rates (short and long compared)

	Jazz		Rock		Speech		VoiceOver	
	Short	**Long**	**Short**	**Long**	**Short**	**Long**	**Short**	**Long**
iBiquity	.9	.5	1.2	1.2	1.4	1.5	1.5	1.4
Orban	.7	.6	1.0	1.1	1.5	1.6	1.5	1.5
VoiceAge	1.2	.9	1.9	1.5	.9	1.2	1.4	1.3
QEI	1.2	1.4	1.5	1.4	2.1	1.8	1.7	1.7
McMartin Light	1.7	2.0	2.4	2.3	2.3	3.0*	2.3	1.9
McMartin Med	2.4	2.7	2.5	3.0*	3.3	3.8*	3.0	2.9
Source	.6	.5	.9	.9	.8	.8	1.2	1.0

*Significantly more impairment heard

Table 4.7.4 - Impairments Reported at Very Low Bit Rates (short and long compared)

	Classical		Jazz		Speech		VoiceOver	
	Short	**Long**	**Short**	**Long**	**Short**	**Long**	**Short**	**Long**
iBiquity	1.3	1.4	1.5	1.7	3.0	3.4*	2.5	2.6
VoiceAge	.4	.9*	1.9	1.6	1.3	1.0	2.0	1.9
Anonymous	1.3	1.2	2.6	3.3*	2.4	2.2	3.1	3.6*
QEI	1.4	1.4	1.2	1.4	2.1	1.8	1.7	1.7
McMartin Light	2.1	2.5	1.7	2.0	2.3	3.0*	2.3	1.9
McMartin Med	3.0	3.2	2.4	2.7	3.3	3.8*	3.0	2.9
Source	.6	.7	.6	.5	.8	.8	1.2	1.0

*Significantly more impairment heard

4.8 Results – Satisfied with the quality of the transmission

Participants were asked whether they would continue to listen to the transmission demonstrating satisfaction with the quality, or whether they would turn off the radio, demonstrating dissatisfaction with the quality. Table 4.8.1 shows participant satisfaction in low bit rates; Table 4.8.2 shows participant satisfaction in very low bit rates. Corroborating quality opinion scores, these percentages indicate that for low bit rates, significantly more listeners would leave their radio on when listening to digitally coded transmissions than when listening to existing SCA transmissions. The only exception to this was in the Rock genre, where approximately the same percentage of people claimed they would listen to QEI and VoiceAge. With regard to very low bit rates, a majority of participants claimed that they would keep VoiceAge on when listening to speech and classical. However, in almost all other cases, QEI was either equally rated or fared better than did the digital coders.

Table 4.8.1 - Percentage of participants willing to leave radio on: Low Bit Rate

	Classical	Jazz	Rock	Speech	VoiceOver
iBiquity	.80*	.90*	.82*	.78*	.76
Orban	.92*	.92*	.76*	.71*	.70
VoiceAge	.88*	.78	.56	.85*	.78
QEI	.72	.77	.62	.54	.70
McMartin Light	.57**	.57**	.48**	.51	.53**
McMartin Med	.35**	.41**	.38**	.23**	.30**
Source	1.00	.93	.88	.93	.81

*Significantly more satisfied than QEI
** Significantly less satisfied than QEI

Table 4.8.2 - Percentage of participants willing to leave radio on: Very Low Bit Rate

	Classical	Jazz	Rock	Speech	VoiceOver
iBiquity	.82*	.68	.44**	.30**	.34**
VoiceAge	.92*	.56**	.36**	.79*	.52**
Anonymous	.78	.38**	.16**	.47	.28**
QEI	.72	.77	.62	.54	.70
McMartin Light	.57	.57	.48	.51	.53
McMartin Med	.35	.41	.38	.23	.30
Source	1.00	.93	.88	.93	.81

*Significantly more satisfied than QEI
** Significantly less satisfied than QEI

Table 4.8.3 and Table 4.8.4 compare short and long listening tests to see if participants changed their opinions after listening for longer periods of time. With regard to low bit rates, in most conditions involving digitally encoded samples, ratings stayed constant or rose slightly. With regard to very low bit rates, in four cases, ratings dropped significantly. Three of these (McMartin Light, McMartin Medium and iBiquity) were in very low bit rate speech, and one (Anonymous) was in very low bit rate Voiceover. All other ratings showed a level of consistency, again suggesting that unless impairments were obvious and annoying, people did not change their minds after listening for long periods.

Table 4.8.3 - Comparison between short and long samples: Low Bit Rate

	Jazz		**Rock**		**Speech**		**Voiceover**	
	Short	**Long**	**Short**	**Long**	**Short**	**Long**	**Short**	**Long**
iBiquity	.90	.96	.82	.88	.78	.80	.76	.84
Orban	.92	.96	.76	.78	.71	.76	.70	.80
VoiceAge	.78	.90	.56	.69	.85	.84	.78	.82
QEI	.77	.76	.62	61	.54	.60	.70	.73
McMartinLight	.57	.59	.48	.59	.51	.34*	.53	.65
McMartinMed	.35	.59	.38	.35	.23	.14*	.30	.30
Source	.93	.99	.88	.90	.93	.95	.81	.89

*Significantly less satisfied after listening to long sample

Table 4.8.4 - Comparison between short and long samples: Very Low Bit Rate

	Classical		**Jazz**		**Speech**		**Voiceover**	
	Short	**Long**	**Short**	**Long**	**Short**	**Long**	**Short**	**Long**
iBiquity	.82	.74	.68	.71	.30	.14*	.34	.35
VoiceAge	.92	.90	.56	.59	.79	.89	.52	.57
Anonymous	.78	.84	.38	.41	.47	.49	.28	.06*
QEI	.72	.86	.77	.76	.54	.60	.70	.73
McMartinLight	.57	.55	.57	.59	.51	.34*	.53	.65
McMartinMed	.35	.24	.41	.45	.23	.14*	.30	.30
Source	1.00	.96	.93	.99	.93	.96	.81	.89

*Significantly less satisfied after listening to long sample

5 Conclusions

General

1. There were no statistical differences between visually impaired and non-visually impaired listeners' quality ratings.
2. There were age and gender differences among participants' ratings, following general trends seen in other audio testing. Older males (above 60) gave all samples higher scores than all other groups, suggesting that they were slightly less discriminating than younger listeners and female listeners.
3. Participants heard differences among coders and conditions, but with few exceptions they rated matching "short" and "long" samples similarly. This was true for quality ratings, impairments heard and satisfaction scores.

Low Bit Rate

1. All of the digital coders that were consumer tested demonstrated upgraded quality to existing SCA receivers, including QEI processed with no impairments, McMartin processed with minor impairments, and McMartin processed with major impairments.
2. Participants heard significantly less impairments in digitally coded recordings than they did in unimpaired SCA recordings. This was surprising, and suggests that people were more sensitive to analog artifacts (e.g., bandwidth limitations, noise) than they were to digital artifacts.
3. Quality ratings, impairment scores and satisfaction scores for digitally encoded recordings did not change when participants listened to longer samples of music.

Very Low Bit Rate

1. Of all the digital coders tested, only VoiceAge demonstrated upgraded quality to existing SCA receivers. This was particularly apparent in speech.
2. iBiquity and Anonymous rated significantly lower than SCA receivers in speech, suggesting that consumers would be less satisfied with digital service than their present service if these coders were selected.
3. Participants heard more impairment in digitally coded samples when compared to QEI, except for VoiceAge in speech.
4. Quality ratings, impairment scores and satisfaction scores for VoiceAge recordings did not change when participants listened to longer samples of music.

6 Acknowledgements

Special thanks goes to the National Public Radio Staff, including John Kean, Jan Andrews, Kyle Evans and Sunny Khemlani for their extensive efforts in all aspects of the test program. Without this staff, the testing would not have been completed. Additionally, thanks to David Noble for recruiting participants and facilitating production of source material. Finally, thanks to all participating coder vendors, VIPs and NVIPs for spending their time and effort to help determine

whether digital coders in the new digital radio system will benefit radio reading service consumers.

Appendix 1

REQUEST FOR INFORMATION

Low Bit Rate Coder for Radio Reading Service Use in the HD Radio® System

Requesting Organizations

National Public Radio
International Association of Audio Information Services
iBiquity Digital
Kenwood USA

Background

The International Association of Audio Information Services (IAAIS) represents audio information services, including radio-reading services, which broadcast timely information to consumers who are visually impaired. In the United States, many of these services are affiliated with NPR member stations and most services are delivered over Subsidiary Communications Authorizations (SCAs), which typically operate on the 67 kHz or 92 kHz subcarrier of FM broadcast stations. Special receivers are provided by the services for local users to receive these broadcasts. Such services have been in operation for more than thirty years and reach an estimated audience of more than a million print disabled consumers across America.

iBiquity Digital is the owner of HD Radio® technology which has been authorized by the U.S. Federal Communications Commission for use by all FM and AM radio stations to provide digital audio transmissions.[2] The FM system utilizes roughly 100 kbps to transmit main audio program material. The FM system design also permits the use of additional bandwidth, known as the Extended Hybrid Partitions to add further data capacity. Each of the four partitions provides approximately12.5 kbps of real time throughput. The FCC is expected to authorize use of the Extended Hybrid Partitions in the near future.

Kenwood USA is a worldwide leader in the manufacture of radio receivers and has made a substantial commitment to developing digital radio services for use by the American people. Kenwood USA and NPR, in close cooperative with iBiquity Digital, have a track record of demonstrating new digital radio service features in the Tomorrow Radio Project and have joined with the IAAIS in supporting the activities identified in this RFI.

In preparation for the eventual use of Extended Hybrid Partitions, the requesting organizations seek information that would support a perceptual test of digital audio coders at bit-rates supported by one or two of these partitions. This RFI invites your company to submit audio samples for consideration in the perceptual tests. The enclosed audio CD contains a series of program audio samples to transcode through your company's coder. Details on preparation of

[2] See the Report and Order delivered in MM 99-325, dated October 10, 2002.

the transcoded results is described below. This RFI also provides a series of issues for written response by your company regarding its coder and its usage. However, these responses need not be returned with the transcoded audio samples – a timetable is provided below.

Technical references of the iBiquity FM Digital Hybrid system can be found at
www.ibiquitydigital.com
www.nab.org/SciTech
www.fcc.gov

Information Sought

Recent advances in low bit rate audio coding technology warrants a service provider and technology stakeholder study to determine the suitability of low bit rate audio coders for use by reading services. Information from the study will potentially aid the migration of radio reading services from FM analog to digital transmission using iBiquity's technology. Audio coding technology will be tested in two categories: (a) coders providing service at 11.8 kbps or below; and (b) coders providing service between 11.9 and 24.2 kbps.

Evaluation Criteria and Operational Objectives

The objective of this RFI is to determine whether any low bit rate audio system(s) have advanced to the point where they can either

- Achieve improved audio quality within a single 12.5 kbps hybrid partition, relative to current analog transmission techniques, as measured by listener controlled testing using MOS rankings (see Specific Procedure for Responding to the Audio Quality Assessment Inquiry, below), or
- Achieve improved audio quality within two 12.5 kbps hybrid partitions, relative to current analog transmission techniques, as measured by listener controlled testing using MOS rankings (see Specific Procedure for Responding to the Audio Quality Assessment Inquiry, below).

As secondary objectives, the study will determine if any low bit-rate systems possess the ability to be upgraded via broadcast delivered code without user intervention,

- Possess the flexibility, innovation and extensibility for providing service enhancements for system integration, and
- Suitability for radio reading services in terms of
    - Ease of integration with the HD Radio protocol and technology implementation into future mass market FM HD Radio receiver chipsets,
    - Fee-free or minimal licensing costs,
    - Additional feature capabilities, such as storage of data streams for on-demand playback, and
    - Stability during annual re-authorizations wherein updates may be downloaded to the system.

Specific Procedure for Responding to the Audio Quality Assessment Inquiry

Respondents to this RFI should use the enclosed audio CD, which has been specifically prepared by NPR and radio reading services. The CD contains samples of speech and voice-over programming generated at the studios of IAAIS members and music selections, including rock, jazz, and classical.

Respondents are welcome to address questions to this RFI or requests for clarification to the requesting organizations at the Email addresses and deadlines provided below. Replies will be provided to all participating respondents without identifying the party initiating the question.

Respondents must return a standard audio CD (44.1 kHz linear PCM, monophonic) containing a transcoded (encoded/decoded) copy of all audio samples provided in Section 2 to demonstrate the overall performance of their audio coder technology. Respondents are free to determine the best parameters for encoding and decoding consistent with a maximum payload of 11.8 kbps or 24.2 kbps and the Operational Objectives outlined above. However, audio parameters and pre-processing of the samples prior to encoding (if any) must be the same for all program samples. If pre-processing is performed external to the coder, respondents must provide a description of the processing technique. Respondents will be categorized according to bit rate, for example, if two respondents submit coded material at 8 kbps and 11 kbps, they would be judged on the same merits. Respondents are encouraged to provide CD recordings of audio transcoded at both bit rates if appropriate for their coder.

Depending on the number of respondents, and in the interest of managing test complexity and time required, the requesting organizations may perform listener audio quality testing with those coders which in their sole discretion, best meet the Evaluation Criteria and Operational Objectives. If the requesting organizations deem it appropriate, respondents should be willing to provide real-time demonstrations of their technology at the studios of NPR in Washington for further evaluation.

Timeframe

Acknowledgement of the intention to participate must be provided to the Project Owners by June 25, 2004. Participants must furnish the audio sample to National Public Radio at the address below, followed by written responses to the Evaluation Criteria and Operational Objectives section.

A brief summary of project milestones includes the following:

29 June 2004 – Deadline for questions relating to this RFI
2 July 2004 – Deadline for participation in the RFI testing by respondents
7 July 2004 – Deliverable deadline – Includes CD with coder-processed audio
30 July 2004 – Deadline for written responses from respondents

Project Owners

David Noble
IAAIS Government Relations
Sun Sounds of Arizona
2323 W 14th Street, Tempe, AZ 85281-6950
Email: david.noble@sunsounds.org

James Paluzzi
IAAIS Technology Committee
Email: jpaluzzi@boisestate.edu

Electronic copies should be sent to both Project Owners as well as one copy to the following representatives of the Requesting Organizations:

John Kean
Senior Technologist, Engineering & Operations
National Public Radio
635 Massachusetts Ave., NW, Washington, D.C. 20001
Email: jkean@npr.org

Mike Bergman
Director of R&D/Digital Broadcast
Kenwood USA
48 Reed Drive South, West Windsor, NJ 08550-2013
Email: mbergman@kenwoodusa.com

Ashruf El-Dinary
IBiquity Digital
Suite 200
8865 Stanford Blvd., Columbia, MD
Email: eldinary@ibiquity.com

RFI Limitations

The issuance of this RFI by the Requesting Organizations neither implies nor constitutes any commitment to issue an RFP or enter into any contractual or other obligation to any person. Those submitting responses do so at their own risk and expense. The Requesting Organizations may withdraw, modify, amend or further pursue the information gathering identified in this RFI at any time without further notice. The Requesting Organizations may seek additional information from any person or information supplier after the due date. No legally enforceable rights shall exist with respect to the subject of this RFI. Any response to this RFI, irrespective of whether a formal response, shall be deemed acceptance to these limitations.

Appendix 2

Names and Affiliations of Expert Listeners

Name of Expert	Affiliation
David Wilson	Consumer Electronics Association
David Layer	National Association of Broadcasters
Mike Starling	National Public Radio
Meghan Hayes	Consumer Electronics Association
Mike Bergman	Kenwood
Richard Cassidy	WAMU
Sean Wallace	Wavetech Technologies
Scott Bridgewater	National Public Radio

Appendix 3

Experimenter's Script – Phase 1

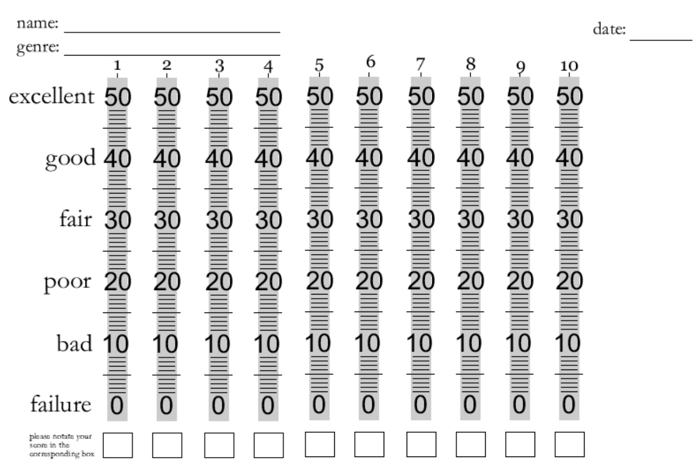
Welcome to our session! Today, you will be participating in a listening experiment, which should last about 2 hours. You will be listening to speech, music and voice-over samples over one loudspeaker. You are going to help us identify 6 coders that we will be using in our Reading Services consumer testing. The procedure you will be using will allow us to rank-order the coders that have been submitted by various companies. We have hidden the names of these companies, as well as the bit-rate. Although you may be curious about what you are listening to, we ask you to refrain from trying to "second guess" this process. We merely want you to tell us your impressions about what you hear.

There will be 14 trials in this test. You will be presented with a play-list through Windows Media Player. We ask you to listen to all of the sound samples once through, and then replay any that you need to in order to make final judgments. You will be ranking these coders according to the scale presented on the sheet. Although it may be difficult, we ask you to rank order the coders – by that I mean give each coder a "unique" score. Some of them will be very similar, and you may want to give them identical scores. Or, they will be very different in "nature", but they may be equally as good or bad, and you will want to give them identical scores. We would like you to try your best to give them different scores – even if those scores are close (for example a 39 and a 40).

Please write the answers on your answer sheet in two ways. First circle the placement on the "bar", then write your numeric answer in the box below.

Let's look at the samples and sheet now, so you can familiarize yourself with the process.

Picture of MUSHRA answer sheet



Appendix 4

Experimenter's Script – Phase 2

Welcome to our session! Today, you will be participating in two listening experiments, which should each last about 2 hours. You will be listening to music and speech samples over one speaker. There are three parts to this study. The first part is a short discrimination test, the second, third and fourth parts are opinion tests. In the opinion tests, there are no right or wrong answers – we are merely seeking your opinion about how different systems sound. You will be working with your panel all day. Since three people are testing at once, please be patient with your fellow participants, especially if they need to hear something again, or take more time with their answers than you might. You will be given plenty of breaks during the listening sessions. If you need some additional break time between samples, please let me know. Please turn off your cell phones, pagers and all other electronic equipment now.

During the tests we are going to ask you lots of questions. Because we don't want you to influence your listening partners with your responses, we are going to ask you to hold up your hand, close to your chest, and tell us what you think by raising a number of fingers on your hand. We will tell you how many fingers to hold up to register your responses to various questions as the tests proceed.

OK, we are ready to examine your ability to hear differences. Let me first explain the task you will be doing, and then we'll start. In this task, your job is to decide which sample (either A or B) is different from the reference. I will play the reference first, then sample A and then sample B. Once you have decided, I will ask you to register your response. Please hold up 1 finger if you think "A" is different, and hold up 2 fingers if you think "B" is different. Please hold up your fingers until I say OK, so I can make sure I record your answers properly. You are welcomed to ask me to play the samples as many times as necessary to make your decisions. The differences between some samples can be small, so please listen very closely. One of the samples *will always be identical to the reference*. The other *will always be different from the reference*.

I will now adjust the listening level for everyone. I won't adjust it during the test again, so make sure this level is comfortable before you start! Hopefully, we will all agree on a reasonable listening level.

Any questions?

Experimenter's Script - The ACR-MOS Test

In this part of the study, you are going to hear many samples. For each sample you hear, you are going to be asked 3 things. First you will be asked to give your opinion of the "overall quality" of the sound samples you hear. You will be presented with one sound sample at a time. Listen carefully to the sample, and then rate the sample on an Excellent to Bad scale. The categories you can choose from are: Excellent, Good, Fair, Poor and Bad. If you want excellent, raise 5 fingers, good, raise 4, fair raise 3, poor raise 2 and bad raise 1. Don't worry about remembering all of this now, we will continually remind you of the numbering scheme throughout the test. 2^{nd}, we will ask you whether you heard impairments on the audio (these are odd sounds, annoying artifacts, things that just sound strange). If you heard something annoying we want you to let us know. The way you do this is to answer the annoyance question: On a 0 – 5 scale, the question is: how annoying were the artifacts? Raise 5 fingers if they were extremely annoying, 4 very annoying, 3 annoying, 2 slightly annoying, 1 heard but not annoying. If you didn't hear any artifacts, just make a fist. Finally, we'd like you to tell us whether you'd keep listening to this audio over your radio, or if you would turn it off because of the quality. For this question, pretend you are reasonably interested in what is being said, but not desperate to hear the program. We all know we'd settle for poor audio if we really want to hear something – but in this test we want you to think about an "average" listening situation in which you are interested, but not "wild" to hear what is being transmitted. Raise 1 finger if you would keep the radio on, and make a fist if you'd turn it off. After around 100 samples, you will be given a 10-15 minute break.

This test is different from the first test you took. There is no stated reference against which to compare the samples you are hearing. You simply hear a recording then rate it. You may probably feel a little unsure of yourself on the first 3 or 4 trials. Don't worry, after the first 3 or 4 samples, you'll feel like a pro!
Don't be afraid to use the entire scales to rate the samples. If you believe the sample sounded excellent, say so! If it sounded bad, again, say so.

Any questions? Great. Now, lets try a trial to get used to the procedure and adjust the sound volume once again.

Experimenter's Script - ACR Long Sample Test

In this test, you will be listening to very long samples. There aren't as many of them, but they are quite long. Please listen from start to finish. Please keep an open mind through the entire sample, and rate it on the whole experience. You will be asked the same 3 questions as you had during the short-sample test. To refresh your memory, these are (a) the Quality question; (b) the Annoyance question; and (c) whether you would keep the radio on or turn it off/switch to something else to listen to.

Appendix 5

Scripts used for speech samples

SHORT SAMPLES

Female High C1
Dr. Altura developed a test for the active form, ionized serum magnesium, which helped her and other researchers uncover latent magnesium deficiencies that have been linked to more than a dozen diseases, even though total magnesium levels in the blood are normal.

Female Low C1
Two years ago, Laurie and Ian Craig decided their "bowling alley" had to go. The 8-by-42-foot deck that stretched across the back of their Falls Church Colonial was in tough shape, with rotting wood and loose rails.

Female Medium C1
The boom in research and funding for the biological sciences -- including genetics and molecular biology -- has been matched by a decline in funding for, and interest in, physics and math.

Male High C1
In the background, the shrill din of the male cicadas' mating chorus gave parts of the campus a tropical, almost alien feel.

Male Low C1
In Chicago studies show one in three black male students and more than half of Hispanic male students drop out of school. Many face danger from gangs, lack of family support, and untreated learning disabilities.

Male Med C1
People talk to Ego Le Bow. He has the qualities of any good therapist. He is calm, seems to listen, and is not easily fazed.

LONG SAMPLES

Female High
The millions of Americans taking calcium supplements could be at risk of distorting this balance, even if their calcium supplement contains magnesium.
Few patients have their magnesium level checked, and even if they do, a simple blood test fails to measure the amount of biologically active magnesium, according to studies by Dr. Burton M. Altura and Dr. Bella T. Altura, who study the physiology and pharmacology of magnesium at the State University of New York in Brooklyn.

Dr. Altura developed a test for the active form, ionized serum magnesium, which helped her and other researchers uncover latent magnesium deficiencies that have been linked to more than a dozen diseases, even though total magnesium levels in the blood are normal.
How Magnesium Promotes Health: Still, few doctors are aware of the many health problems that can cause a magnesium deficiency, as well as the role of magnesium in diseases, including heart disease, hypertension, diabetes, asthma, obesity, infertility, migraine, muscle pains, premenstrual syndrome and traumatic stress.

Female Medium

American science is growing weaker, although not across the board. The boom in research and funding for the biological sciences -- including genetics and molecular biology -- has been matched by a decline in funding for, and interest in, physics and math. Because the decline has multiple sources, the solutions will have to be multiple. Poor teaching, and especially poor high school math teaching, bears part of the blame. Even in an era of heavy testing, the standards for high school math are very low. The American Diploma Project pointed out that few states even require the basic Algebra I-Geometry-Algebra II sequence needed just for many entry-level jobs, let alone for higher education. Low expectations, in turn, have led to a dearth of teachers. In March, The Post reported that because of the lack of trained Americans, urban school districts across the country must now rely on international recruitment and generous visa rules to find any high school math and science teachers at all.

Male High

Martin Wikelski has gotten used to having puzzled, lingering glances cast his way as he makes his unusual treks across Princeton University's campus. After all, it's a bit extraordinary for someone to meander outdoors guided by a hand-held 3-foot antenna that resembles a telecommunications relic once commonly placed on rooftops for TV reception. Yesterday, Wikelski went about his task yet again, pointing his six-pronged antenna this way and that as he ambled from one spot to the next in pursuit of his prey, the noisy 17-year cicadas that have recently emerged from underground. All the while, he pushed his content 9-month-old daughter in her stroller. Wikelski patiently listened for the static on his radio receiver to give way to the faint pinging sound that would guide him to his quarry's location. In the background, the shrill din of the male cicadas' mating chorus gave parts of the campus a tropical, almost alien feel.

Male Medium

People talk to Ego Le Bow. He has the qualities of any good therapist. He is calm, seems to listen, and is not easily fazed. He is also a hulking white dog who at times resembles a polar bear. "They whisper in his ear – tell him their dreams," said Dr. Michelle Labow - a psychologist on the Upper West Side who began involving Ego in her sessions with her patients eight years ago.

The story of how Ego became a quasi-psychotherapist is as layered and quirky as any New York tale, but has its roots in this basic fact: it takes creativity to integrate a giant breed dog into one's life in the city. Ego, a Great Pyrenees, hails from a long line of ancestors bred to guard sheep in the mountains.

Appendix 6

Breakdown of ACR-MOS by VIPs and NVIPs

Low Bit Rates

		iBiquity	Orban	VoiceAge	McMartin Light	McMartin Med	QEI
Classical	**VIP**	4.2	4.3	4.2	3.2	2.7	3.5
	NVIP	4.3	4.5	3.8	3.4	2.7	3.8
Jazz	**VIP**	4.3	4.4	4.1	3.3	3.0	3.7
	NVIP	4.1	4.1	3.9	3.5	2.8	3.8
Rock	**VIP**	3.9	4.2	3.5	2.8	2.8	3.4
	NVIP	3.8	4.0	3.3	2.9	2.6	3.3
Speech	**VIP**	3.8	3.8	4.2	2.9	2.2	3.0
	NVIP	3.6	3.5	4.0	3.0	2.2	3.1
Voiceover	**VIP**	3.8	3.7	3.9	3.0	2.4	3.3
	NVIP	3.6	3.4	3.6	3.0	2.3	3.5

Very Low Bit Rates

		iBiquity	VoiceAge	Anonymous	McMartin Light	McMartin Med	QEI
Classical	**VIP**	3.7	4.4	4.0	3.2	2.7	3.5
	NVIP	4.2	4.1	4.0	3.4	2.7	3.8
Jazz	**VIP**	3.6	3.4	2.6	3.3	3.0	3.7
	NVIP	3.6	3.3	3.2	3.5	2.8	3.8
Rock	**VIP**	3.2	2.9	2.1	2.8	2.8	3.4
	NVIP	3.0	3.0	3.0	2.9	2.6	3.3
Speech	**VIP**	2.3	3.8	2.6	2.9	2.2	3.0
	NVIP	2.0	3.8	2.8	3.0	2.2	3.1
Voiceover	**VIP**	2.8	3.3	2.2	3.0	2.4	3.3
	NVIP	2.8	3.3	2.5	3.0	2.3	3.5