**INTERNATIONAL ORGANIZATION FOR STANDARDIZATION**

**ORGANISATION INTERNATIONALE NORMALISATION**

**ISO/IEC JTC 1/SC 29/WG 11**

**CODING OF MOVING PICTURES AND AUDIO**

**WG 11 N14931**

**October 2014, Strasbourg, FR**

|  |  |
| --- | --- |
| **Source** | **Audio Subgroup** |
| **Status** | **Approved** |
| **Title** | **Report on MPEG-H 3D Audio Performance** |

# Introduction

The Call for Proposals for 3D Audio (Call) [1] was issued at the 103rd MPEG meeting held in Geneva, CH in January 2013. Submissions to the Call were evaluated at the 105th MPEG meeting held in Vienna, AT, July 2013. From the 106th to the 109th MPEG meetings, the MPEG-H 3D Audio technology was developed in a collaborative process. In July 2014 at 109th meeting, the standard progressed to DIS, at which point it was technically frozen. It is expected to progress to IS early in 2015.

This document reports performance of MPEG-H 3D Audio, as subjective quality at a number of bit rates. The subjective test data is from the tests associated with the evaluation of Call responses. However, most of the collaborative work pertained to development of metadata structures, and so it is expected that the results of the tests associated with the Call are a fair evaluation of the performance of the issued standard.

# Test Material

The test material was divided into two sets:

* Set 1 was 12 items that were a mix of Channel-based and Object (C/O) signals
* Set 2 was 12 items that were Higher Order Ambisonics (HOA) signals of various orders

The C/O original test items are monophonic WAV files, one for each loudspeaker signal or for each object signal. The items may be for different loudspeaker layouts, e.g. 9.0 to 22.2. When presented in the subjective test, the original test items (i.e. as reference or hidden reference) were played out to the intended loudspeaker layout (i.e. 9.0 original items were played on a 9.0 loudspeaker layout). The loudspeaker layouts for the material and for the subjective test are shown in ANNEX 1.

The HOA original test items were in time-domain coefficient format (see [2]) and were rendered to a 22.2 channel loudspeaker layout for presentation in the tests as reference or hidden reference.

## Channel/Object Material

Table 1 below lists the Channel/Object test items. All items with objects were rendered to loudspeaker signals for a 22.2 channel layout. Since some items have fewer than 22.2 channels, bitrate per item in the tests was adjusted as follows:

item\_bitrate = C\_Factor \* target\_bitrate

For each test at the nominal rates Rate1, Rate2, Rate3, the actual bitrate for each item is shown in the rightmost three columns of the table.

Table 1 - Channel/Object test items.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Label** | **Name** | **Duration (sec)** | **No. Channels** | **No. Objects** | **C\_Factor** | **Rate 1** | **Rate 2** | **Rate 3** |
| CO\_01 | CO\_01\_Church | 16 | 22.2 |  | 1 | 1200 | 512 | 256 |
| CO\_02 | CO\_02\_OMensch | 28 | 22.2 |  | 1 | 1200 | 512 | 256 |
| CO\_03 | CO\_03\_SLNiseko | 19 | 22.2 |  | 1 | 1200 | 512 | 256 |
| CO\_04 | CO\_04\_Fountain\_Music\_3D | 20 | 14 |  | 0.82 | 984 | 420 | 210 |
| CO\_05 | CO\_05\_BarcoTR | 22 | 11.1 |  | 0.75 | 900 | 384 | 192 |
| CO\_06 | CO\_06\_ClarinetConcerto | 24 | 9 |  | 0.7 | 840 | 358 | 179 |
| CO\_07 | CO\_07\_Musikverein1 | 25 | 9 |  | 0.7 | 840 | 358 | 179 |
| CO\_08 | CO\_08\_Musikverein2 | 28 | 9 |  | 0.7 | 840 | 358 | 179 |
| CO\_09 | CO\_09\_Rain | 21 | 22.2 | 1 | 1 | 1200 | 512 | 256 |
| CO\_10 | CO\_10\_Music | 21 | 22.2 | 6 | 1 | 1200 | 512 | 256 |
| CO\_11 | CO\_11\_Betty3 | 10 |  | 38 | 1 | 1200 | 512 | 256 |
| CO\_12 | CO\_12\_Mechanism4 | 9 |  | 31 | 1 | 1200 | 512 | 256 |

## HOA Material

Table 2 below lists the HOA test items. HOA order varied from 3 to 6.

Table 2 - Higher Order Ambisonics (HOA) test items

|  |  |  |  |
| --- | --- | --- | --- |
| **Label** | **Name** | **Duration****(sec)** | **HOA Order** |
| H\_01 | H\_01\_SynthBumbleBee | 20 | 6 |
| H\_02 | H\_02\_Drums1 | 10 | 4 |
| H\_03 | H\_03\_Modern | 12 | 4 |
| H\_04 | H\_04\_Stadium2 | 21 | 4 |
| H\_05 | H\_05\_Water | 20 | 4 |
| H\_06 | H\_06\_Helicopter | 20 | 4 |
| H\_07 | H\_07\_Vocal1 | 20 | 4 |
| H\_08 | H\_08\_BeginningOfAConcert | 20 | 4 |
| H\_09 | H\_09\_ModernElectronicMusic | 20 | 4 |
| H\_10 | H\_10\_Orchestra2 | 20 | 4 |
| H\_11 | H\_11\_ShoutingAudience | 20 | 4 |
| H\_12 | H\_12\_Radio2 | 18 | 3 |

# Test Methodology

All tests used the MUSHRA method described in [1]. A quality scale is used where the intervals are labeled "bad", "poor", "fair", "good" and "excellent". The subjective response is recorded on a scale ranging from 0 to 100, with no decimals digits. The correspondence between labels and score ranges is shown in Table 3 below.

Table 3 - MUSHRA scale

|  |  |
| --- | --- |
| **Descriptor** | **Range of Score** |
| EXCELLENT | Value between 81 and 100 |
| GOOD | Value between 61 and 80 |
| FAIR | Value between 41 and 60 |
| POOR | Value between 21 and 40 |
| BAD | Value between 0 and 20  |

# Systems Under Test

Table 4 below lists the systems in the subjective test. The first is the MPEG-H 3D Audio system (3D Audio). The second and third are the hidden reference (HR), which is the high anchor, and the 3.5 kHz low pass filtered version of the reference (LP35), which is the low anchor. Other responses to the Call are indicated as Anchors A, B, C. Note that the tests for signal set 1 (CO signals) included three additional systems (A, B, C) and the tests for signal set 2 (HOA signals) included two additional systems (A, B)

Table 4 – Systems under test, where 3D Audio is the MPEG-H 3D Audio system and A, B, C are other system that were evaluated in the Call

|  |  |  |  |
| --- | --- | --- | --- |
| **Description** | **ID** | **CO** | **HOA** |
| MPEG-H 3D Audio  | 3D Audio | X | X |
| Hidden Reference | HR | X | X |
| Low Anchor | LP35 | X | X |
| Anchor A | A | X | X |
| Anchor B | B | X | X |
| Anchor C | C | X |  |

# Tests Conducted

The subjective listening test results reported here, collectively referred to as Test1.1 and Test1.3, are a sub-set of those conducted as part of the evaluation of Call responses.

## Loudspeaker Tests

Test1.1 was conducted as six subjective tests: one test for each of three nominal coding bitrates for each of the two signal sets (CO and HOA), with presentation via loudspeakers. As discussed above, that actual coding bitrate for each CO item in each test is shown in Table 1.

## Binauralized Headphone Tests

Test1.3 was conducted as two subjective tests: one test for each of the two signal sets (CO and HOA). The bitrate was Rate 2 (nominal 512 kb/s) for CO signals and 512 kb/s for HOA signals. Two important features of the test are, first, that the Rate 2/512 kb/s bitstream from Test1.1 was used for this test, and second, that the bitstream was decoded and then binauralized for presentation via headphones. This permits the user to have an immersive audio experience via headphones. For reasons of computational efficiency, different binauralization engines were used for CO and HOA signals (i.e. for processing of QM filterbank domain or time domain signals). A binaural room impulse response (BRIR) that was used for this test was recorded in one of the test lab multichannel listening rooms.

A detailed specification of all tests is in [1].

Table 5 – Description of tests conducted

|  |  |  |  |
| --- | --- | --- | --- |
| **Test** | **Nominal Bitrate** | **Test ID** | **Speaker** **Configuration** |
| Test1.1 | 1.2 Mb/s | T1-1\_1200 | 22.2 |
| 512 kb/s | T1-1\_512 | 22.2 |
| 256 kb/s | T1-1\_256 | 22.2 |
| Test1.3 | 512 kb/s | T1-3 | Binauralized stereo (headphones) |

# Listening Tests

The Listening Lab tests sites are shown in the following table.

Table 6 - Listening Labs

|  |  |
| --- | --- |
| **Company** | **ID** |
| ETRI | ETRI |
| FhG-IDMT | IDMT |
| FhG-IIS | IIS |
| Huawei | HUA |
| NHK | NHK |
| Orange | ORL |
| Qualcomm | QUAL |
| Samsung | SAM |
| Sony | SONY |
| Technicolor | TECH |

The following table shows the number of test subjects, after post-screening, from all of the labs that participated in testing. See Section 7 below for a description of the post-screening procedure.

Table 7 - Number of subjects in each test.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Test** | **Nominal Bitrate** | **Test ID** | **CO subjects** | **HOA subjects** |
| Test1.1 | 1.2 Mb/s | T1-1\_1200 | 36 | 28 |
| 512 kb/s | T1-1\_512 | 36 | 26 |
| 256 kb/s | T1-1\_256 | 39 | 25 |
| Test1.3 | 512 kb/s | T1-3 | 46 | 27 |

# Post-screening of Subjective Scores

The listener scores for all tests were post-screened using the following rule: all data associated with subjects satisfying one or both of the following criteria are removed from the corresponding test.

* The subject’s score for the hidden reference stimuli was below 90.
* The subject’s scores for the 3.5 kHz anchor is greater than the hidden reference score.

# Data Analysis

The subjective scores for the CO and HOA signal sets were collected into separate Excel workbooks. Each Excel workbook had a separate sheet for each test’s scores and a Pivot Table associated with the scores.

In addition, an Analysis of Variance (ANOVA) was done on the test scores. The model used was:

Score = Lab + Listener + Signal + System

In this model, the factors were: Lab, Listener, Signal and System, but there were no interactions. The analysis included histograms and normal probability plots of the model residual error, and boxplots of the distribution of the model residual for the factor Lab.

ANOVA results showed a nearly Gaussian distribution of the residual which indicated that parametric statistics were appropriate, and so parametric statistics were used to analyse the data. A nearly equal variance for each Listening Lab indicated that pooling the test scores across all Listening Labs was appropriate, and so this pooling was done for the test.

# Performance Results

## Channel/Object Overall Performance for Test1.1

The following figure shows the mean subjective score, as averages over all test items, for the systems under test at the three tested bitrates. Note that each of the CO signals has its own bitrate, as shown in Table 1. The systems under test are indicated in the legend at the right of the plot. The MUSHRA test methodology requires that the identified hidden reference (HR) receive a score of 100, and this is shown as the blue line at the top of the graph. The MPEG-H 3D Audio system (3D Audio) is the orange line which always has a subjective score that is better than the other systems and low anchor (A, B, C and LP35).

The graph shows that the performance of MPEG-H 3D Audio (3D Audio) ranges from Excellent (Rate 1) to Good (Rate 3). The mean scores are given in Table 9 (see ANNEX 2), along with the 95% confidence interval on the mean.

Figure 1 - Performance for CO signal set

## HOA Overall Performance Test1.1

The following figure shows the mean subjective score, as averages over all test items, for the systems under test at the three tested bitrates (1.2 Mb/s, 512 kb/s and 256 kb/s). The systems under test are indicated in the legend at the right of the plot. The MUSHRA test methodology requires that the identified hidden reference (HR) receive a score of 100, and this is shown as the blue line at the top of the graph. The MPEG-H 3D Audio system (3D Audio) is the sea-green line which always has a subjective score that is better than the other systems and low anchor (A, B and LP35).

The graph shows that the performance of MPEG-H 3D Audio (3D Audio) ranges from Excellent (Rate 1) to Good (Rate 3). The mean scores are given in Table 10 (see ANNEX 2), along with the 95% confidence interval on the mean.

Figure 2 - Performance for HOA signal set

## Channel/Object Overall Performance for Test1.3

The following figure shows the mean subjective score, as averages over all test items, for the systems under test at the Rate 2 bitrate (nominal 512 kb/s). The systems under test are indicated on the horizontal axis. The graph shows that the performance of MPEG-H 3D Audio (3D Audio) is Excellent. The mean scores are given in Table 9 (see ANNEX 2), along with the 95% confidence interval on the mean.

Figure 3 - CO signals binauralized for headphones

## HOA Overall Performance Test1.3

The following figure shows the mean subjective score, as averages over all test items, for the systems under test at 512 kb/s. The systems under test are indicated on the horizontal axis. The graph shows that the performance of MPEG-H 3D Audio (3D Audio) is Excellent. The mean scores are given in Table 9 (see ANNEX 2), along with the 95% confidence interval on the mean.

Figure 4 - HOA Signals binauralized for headphones

# References

All references are available at <http://mpeg.chiariglione.org>

1. N13411, Call for Proposals for 3D Audio
2. N12412, Encoder Input Format for MPEG-H 3D Audio

# ANNEX 1 – Loudspeaker Positions

**Table 8** - Loudspeaker positions and tolerances for the different test setups: M middle, U upper, T top and L lower layer. Output Formats indicate a loudspeaker configuration; Input Formats indicate the loudspeaker positions for playback of channel-based test items.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  | **Output Formats** | **Input Formats** |
| **No.** | **LSLabel** | **Az °** | **Az. Tol. °** | **El. °** | **El.****Tol. °** | **O-5.1**  | **O-8.1** | **O-10.1**  | **O-22.2**  | **I-8.1** | **I-9.1** | **I-11.1** | **I-22.2** |
| **1** | M+000 | 0 | ±2 | 0 | ±2 | X |  | X | X |  | X | X | X |
| **2** | M+030 | 30 | ±2 | 0 | ±2 | X | X | X | X | X | X | X | X |
| **3** | M-030 | -30 | ±2 | 0 | ±2 | X | X | X | X | X | X | X | X |
| **4** | M+060 | 60 | ±2 | 0 | ±2 |  |  |  | X |  |  |  | X |
| **5** | M-060 | -60 | ±2 | 0 | ±2 |  |  |  | X |  |  |  | X |
| **6** | M+090 | 90 | ±5 | 0 | ±2 |  |  |  | X |  |  |  | X |
| **7** | M-090 | -90 | ±5 | 0 | ±2 |  |  |  | X |  |  |  | X |
| **8** | M+110 | 110 | ±5 | 0 | ±2 | X | X | X |  | X | X | X |  |
| **9** | M-110 | -110 | ±5 | 0 | ±2 | X | X | X |  | X | X | X |  |
| **10** | M+135 | 135 | ±5 | 0 | ±2 |  |  |  | X |  |  |  | X |
| **11** | M-135 | -135 | ±5 | 0 | ±2 |  |  |  | X |  |  |  | X |
| **12** | M+180 | 180 | ±5 | 0 | ±2 |  |  |  | X |  |  |  | X |
| **13** | U+000 | 0 | ±2 | 35 | ±10 |  | X |  | X |  |  | X | X |
| **14** | U+045 | 45 | ±5 | 35 | ±10 |  |  |  | X |  |  |  | X |
| **15** | U-045 | -45 | ±5 | 35 | ±10 |  |  |  | X |  |  |  | X |
| **16** | U+030 | 30 | ±5 | 35 | ±10 |  | X | X |  | X | X | X |  |
| **17** | U-030 | -30 | ±5 | 35 | ±10 |  | X | X |  | X | X | X |  |
| **18** | U+090 | 90 | ±5 | 35 | ±10 |  |  |  | X |  |  |  | X |
| **19** | U-090 | -90 | ±5 | 35 | ±10 |  |  |  | X |  |  |  | X |
| **20** | U+110 | 110 | ±5 | 35 | ±10 |  |  | X |  | X | X | X |  |
| **21** | U-110 | -110 | ±5 | 35 | ±10 |  |  | X |  | X | X | X |  |
| **22** | U+135 | 135 | ±5 | 35 | ±10 |  |  |  | X |  |  |  | X |
| **23** | U-135 | -135 | ±5 | 35 | ±10 |  |  |  | X |  |  |  | X |
| **24** | U+180 | 180 | ±5 | 35 | ±10 |  |  |  | X |  |  |  | X |
| **25** | T+000 | 0 | ±2 | 90 | ±10 |  |  | X | X |  |  | X | X |
| **26** | L+000 | 0 | ±2 | -15 | +5-25 |  | X |  | X |  |  |  | X |
| **27** | L+045 | 45 | ±5 | -15 | +5-25 |  |  |  | X |  |  |  | X |
| **28** | L-045 | -45 | ±5 | -15 | +5-25 |  |  |  | X |  |  |  | X |
| **29** | LFE1 |  45 | ±15 | -15 | ±15 | X  | X | X | X | X | X | X | X |
| **30** | LFE2 | -45 | ±15 | -15 | ±15 |  |  |  | X |  |  |  | X |

# ANNEX 2 – Test Scores

## Channel/Object Signal Set

Table 9 below shows the scores for the CO signal set for all three bit rates. The table shows mean scores, as pooled over all test items, and 95% confidence intervals. The column heading High shows the upper point and Low shows the lower point of the confidence interval. Heading CI shows half of the confidence interval length.

Table 9 - Scores for CO signal set

|  |  |  |  |
| --- | --- | --- | --- |
| **CO\_Test1-1\_1200** |  |  |  |
|  |  |  |  |  |
| CI | Sys | High | Low | Mean |
| 0.17 | HR | 99.54 | 99.20 | 99.37 |
| 0.94 | LP35 | 14.42 | 12.54 | 13.48 |
| 0.85 | A | 94.44 | 92.74 | 93.59 |
| 0.64 | B | 96.46 | 95.18 | 95.82 |
| 0.46 | C | 97.92 | 97.00 | 97.46 |
| 0.49 | 3DAudio | 97.56 | 96.58 | 97.07 |
|  |  |  |  |  |
| **CO\_Test1-1\_512** |  |  |  |
|  |  |  |  |  |
| CI | Sys | High | Low | Mean |
| 0.10 | HR | 99.90 | 99.69 | 99.79 |
| 0.88 | LP35 | 14.56 | 12.80 | 13.68 |
| 1.80 | A | 75.44 | 71.83 | 73.63 |
| 1.59 | B | 78.34 | 75.16 | 76.75 |
| 1.66 | C | 80.69 | 77.37 | 79.03 |
| 1.29 | 3DAudio | 86.25 | 83.68 | 84.96 |
|  |  |  |  |  |
| **CO\_Test1-1\_256** |  |  |  |
|  |  |  |  |  |
| CI | Sys | High | Low | Mean |
| 0.10 | HR | 99.91 | 99.72 | 99.82 |
| 0.83 | LP35 | 14.18 | 12.53 | 13.36 |
| 1.86 | A | 68.09 | 64.36 | 66.22 |
| 1.79 | B | 54.62 | 51.05 | 52.84 |
| 1.55 | C | 73.91 | 70.80 | 72.36 |
| 1.58 | 3DAudio | 74.77 | 71.61 | 73.19 |

The information in the tables above is shown graphically in the plots below. The corresponding plot for the headphone test is found in Section 9.3.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **CO\_Test13** |  |  |  |  |
|  |  |  |  |  |
| CI | Sys | High | Low | Mean |
| 0.12 | HR | 99.84 | 99.59 | 99.71 |
| 1.05 | LP35 | 20.77 | 18.67 | 19.72 |
| 1.04 | 3DAudio | 89.67 | 87.60 | 88.64 |
| 1.53 | A | 80.19 | 77.13 | 78.66 |
| 1.36 | B | 82.99 | 80.27 | 81.63 |
| 1.82 | C | 73.55 | 69.91 | 71.73 |

## HOA Signal Set

Table 10 below shows the scores for the HOA signal set for all three bit rates. The table shows mean scores, as pooled over all test items, and 95% confidence intervals. The column heading High shows the upper point and Low shows the lower point of the confidence interval. Heading CI shows half of the confidence interval length.

Table 10 - Scores for HOA signal set

|  |  |  |  |
| --- | --- | --- | --- |
| **HOA\_Test1-1\_1200** |  |  |  |
|  |  |  |  |  |
| CI | Sys | High | Low | Mean |
| 0.29 | HR | 98.74 | 98.15 | 98.45 |
| 1.31 | LP35 | 22.34 | 19.72 | 21.03 |
| 1.05 | A | 92.91 | 90.80 | 91.85 |
| 0.76 | B | 96.15 | 94.63 | 95.39 |
| 0.73 | 3DAudio | 96.17 | 94.71 | 95.44 |
|  |  |  |  |  |
| **HOA\_Test1-1\_512** |  |  |  |
|  |  |  |  |  |
| CI | Sys | High | Low | Mean |
| 0.21 | HR | 99.55 | 99.14 | 99.35 |
| 1.31 | LP35 | 21.45 | 18.83 | 20.14 |
| 1.61 | A | 86.72 | 83.50 | 85.11 |
| 1.30 | B | 88.73 | 86.13 | 87.43 |
| 1.26 | 3DAudio | 88.97 | 86.44 | 87.71 |
|  |  |  |  |  |
| **HOA\_Test1-1\_256** |  |  |  |
|  |  |  |  |  |
| CI | Sys | High | Low | Mean |
| 0.16 | HR | 99.90 | 99.58 | 99.74 |
| 1.18 | LP35 | 21.02 | 18.66 | 19.84 |
| 2.22 | A | 75.40 | 70.97 | 73.18 |
| 1.98 | B | 75.55 | 71.59 | 73.57 |
| 1.93 | 3DAudio | 76.62 | 72.75 | 74.69 |

|  |  |  |  |
| --- | --- | --- | --- |
| **HOA\_Test1-3** |  |  |  |
|  |  |  |  |  |
| CI | Sys | High | Low | Mean |
| 0.35 | HR | 99.82 | 99.11 | 99.46 |
| 1.72 | LP35 | 11.30 | 7.87 | 9.58 |
| 2.09 | Sys1 | 86.14 | 81.96 | 84.05 |
| 2.10 | Sys2 | 90.05 | 85.84 | 87.94 |
| 1.98 | Sys3 | 89.94 | 85.97 | 87.95 |

The information in the tables above is shown graphically in the plots below. The corresponding plot for the headphone test is found in Section 9.4.