|  |  |
| --- | --- |
| **Joint Collaborative Team on Video Coding (JCT-VC)**  **of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11**  24th Meeting: Geneva, CH, 26 May – 1 June 2016 | Document: JCTVC-X0080r1 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **Supplement to the SHVC verification test** | | |
| *Status:* | Input Document from JCT-VC | | |
| *Purpose:* | Information | | |
| *Author(s) or Contact(s):* | V. Baroncini (MPEG Test Chair) | Email: | [baroncini@gmx.com](mailto:baroncini@gmx.com) |
| *Source:* | JCT-VC - Test Chair | | |

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Introduction

This document contains the results of supplement the SHVC verification tests. In the SHVC verification tests [7], one of the test cases compared the *performance of SHVC versus HEVC simulcast* when the enhancement layer content is HDR content coded in BT.2020 and base layer content is coded in BT.709.

The supplemental verification test evaluates the performances of the dual-layer coder (with both the SDR base layer and the HDR enhancement layers are coded in BT.2020) against the base layer coder. The base layer conforms to HEVC Main 10 coding as specified in HEVC version 2. The draft plan for the supplementation verification was described in [9]; testing conditions and video material preparation are again summarized in this document. The coding performance of the SDR to HDR scalability is tested using the Scalable Main 10 profile.

Subjective tests were conducted from 11 to 22 of May 2016. The tests compared the scalable solution against the HDR encoding configuration for HEVC Main 10 coding described in JCTVC-W1017 [1] and as implemented in JCTVC-W1020 [5]. The SHVC coding uses the SHM software as described in JCTVC-X0035 [8].

The bitrate savings (based on MOS BD-rate) of SHVC solution when compared to simulcast are reported to be on an average 45%, and the bitrate overhead (based on MOS BD-rate) of the SHVC solution when compared to single layer HEVC solution is reported to be on an average 15%, when both the base layer and enhancement layers are using BT.2020.

# Schedule

The following schedule was applied for the subjective test:

* Bitstreams encoded and made available by May 11, 2016
* Subjective evaluation starts: May 11, 2016
* Subjective evaluation completed: May 22, 2016
* Subjective evaluation results available: May 24, 2016

# Test sites

The subjective test shall be carried out at the following test sites:

* GBTech Laboratory, Italy
* Waitek Laboratory, Italy

# Test conditions

The single-layer HDR was encoded using tools described in JCTVC-W1017 [1] and as implemented in JCTVC-W1020 [5]. The SHVC coding uses the SHM software as described in JCTVC-X0035 [8], where the tools described in [1] were applied for the enhancement layer coding.

The SHVC and HEVC simulcast performance were compared by encoding corresponding sequences at selected constant QP. The QP settings for SHVC were selected to keep the SHVC bitrate ratio between enhancement layer and the base layer at around 10%-20%. For simulcast, QP value for the SDR layer are chosen to be the same as used for the base layer of SHVC, and the QP value of the decoded HDR layer are chosen such that the quality of the HDR layer is roughly similar subjective quality as the decoded EL SDR.

The following test conditions were used for the verification test:

1. Number of sequences, resolutions, and frame rates
   1. 7 sequences in BT.2020 for base and enhancement layers
   2. 4 sequences are 24 or 25 fps and 3 sequences are 50 fps
2. Bitstreams
   1. Anchor 3.2 configuration bitstreams generated as described in Annex B
   2. SHM configuration bitstreams generated as described in Annex B
3. Encoding parameters
   1. Quantization
      1. 4 bit rate points per sequence
   2. Bit depth
      1. 10 bit encoding for all sequences
   3. Coding structure
      1. Random access coding structure using 7 B-pictures in a hierarchical sub-GOP structure
      2. IRAP picture period at approximately 1.0 second
   4. Other settings as in the configuration files as described in Annex B

# Test Sequences

The test sequences used in the tests are listed in Table 1. Single layer HDR bitstreams were generated using HEVC Main 10 Profile with encoder settings as in [1], BT.2020 container, ST 2084 Transfer Function and NCL Y’CbCr colour space conversion. Dual layer bitstreams were generated using SHM Main 10 Profile, BT.2020 container, BT.709 Transfer function for base layer, ST 2084 Transfer Function for enhancement layer and NCL Y’CbCr colour space conversion.

All the test sequences have the following characteristics:

* Container: BT.2020 (Gamut: BT. 709 or P3D65 depending on the content; see table 1)
* Original (not coding) colour format: RGB 4:4:4
* Coding format: 10 bit 4:2:0

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sequence | Frame rate | Gamut | Length (frames) | Length (seconds) |
| Market3 | 50 | BT.709 | 400 | 8 |
| Showgirl | 25 | P3D65 | 339 | 13,56 |
| EBU\_06\_Starting | 50 | BT.709 | 500 | 10 |
| EBU\_04\_Hurdles | 50 | BT.709 | 500 | 10 |
| WarmNight | 24 | P3D65 | 361 | 15 |
| BalloonFestival | 24 | BT.709 | 240 | 10 |
| GarageExit | 24 | P3D65 | 288 | 12 |

Table 1: Selected test sequences

The EBU\_04\_Hurdles and EBU\_06\_Starting are originally shot in 100 fps and were played out in 50 fps (slow motion), which follows the visual testing practice adopted by HDR AhG of MPEG and JCTVC. The SDR sequences for Market3, Showgirl, and WarmNight are manually graded and the SDR sequences for BalloonFestival, GarageExit, EBU\_04\_Hurdles and EBU\_06\_Starting are automatically graded. Six out of the seven sequences were tested. The colour gamut of Market3, EBU\_06\_Starting, EBU\_04\_Hurdles, BalloonFestival and GarageExit is the same for the base and enhancement layers. For the WarmNight and ShowGirl sequences, the colour gamut of the base layer sequences is BT.709.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Sequence | Single Layer | | | | Dual Layer | | | | | | | |
| QP1 | QP2 | QP3 | QP4 | QP1 | | QP2 | | QP3 | | QP4 | |
| BL | EL | BL | EL | BL | EL | BL | EL |
| Market3 | 21 | 28 | 31 | 33 | 22 | 22 | 29 | 29 | 32 | 32 | 34 | 34 |
| Showgirl | 15 | 19 | 23 | 27 | 17 | 19 | 22 | 24 | 26 | 28 | 30 | 32 |
| EBU\_06\_Starting | 20 | 23 | 29 | 33 | 21 | 22 | 24 | 26 | 29 | 32 | 33 | 36 |
| EBU\_04\_Hurdles | 21 | 25 | 28 | 32 | 23 | 23 | 27 | 27 | 30 | 32 | 34 | 36 |
| WarmNight | 14 | 19 | 23 | 27 | 18 | 18 | 23 | 23 | 27 | 27 | 31 | 31 |
| BalloonFestival | 22 | 27 | 30 | 32 | 22 | 22 | 27 | 26 | 30 | 29 | 32 | 31 |
| GarageExit | 21 | 25 | 30 | 34 | 21 | 21 | 25 | 25 | 30 | 30 | 34 | 34 |

Table 2: Selected single layer dual layer QPs

# Test design

The tests included the following cases resulting by the combination of:

# 7 sequences,

# 4 compression rates,

# 2 coding conditions.

Following the above conditions a total of 56 Processed Video Sequences (PVS) were produced and evaluated in the test. The file format was AVI in LogLuv format, in a way the SIM2 HDR display could show them in a proper HDR fashion.

The test method DCR Variant II (*A B A\* B\* Vote*), as defined in Recommendation P.910 [6] was used.

Assuming that a test sequence is 10 seconds long, the DCR Variant II Basic Test Cell (see figure 1 of Annex A) is long 47 seconds; in our case the length of the test sequences were different as shown in Table 1. This leads to an actual mean length of the test sequences of 11,2 seconds and consequently a mean length of the BTCs equal to 52 seconds.

To evaluate the 56 PVSs the test sessions were designed made of:

# stabilization phase, made of 3 BTC;

# evaluation of PVS, made of 14 BTC;

# original vs. original, made of 2 BTC.

This leads to a total of four sessions, with a length of around 16 minutes and 30 seconds each.

In subjective viewing sessions, the higher quality video corresponding to the enhancement layer video was displayed as the “coded” video, the lower quality video corresponding to the base layer video was not viewed. Two test cases are compared: the HEVC simulcast coded higher quality video (Case A, to be labelled as P01), and the SHVC coded higher quality video (Case B, to be labelled as P02).

The sequences are renamed as “P0xS0xCXRn” using the following convention:

* P0x identifies the test case, with Case A being HEVC simulcast coded and Case B being SHVC coded.
* S0x identifies the sequence number, from S01 to S07.
* Rn identifies the coded rate, with four rates R1 to R4 identifying rates from the highest to the lowest.

Table 4‑2 Numbering of test sequences according to viewing session

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sequence no | Sequence name | Frame rate | Frame rate to be displayed | Coding Config. |
| P0xS01C4Rn | Market | 50 fps | 50 fps | RA |
| P0xS02C4Rn | ShowGirl | 25 fps | 25 fps | RA |
| P0xS03C4Rn | EBU\_06\_Starting | 50 fps | 50 fps | RA |
| P0xS04C4Rn | EBU\_04\_Hurdles | 50 fps | 50 fps | RA |
| P0xS05C1Rn | GarageExit | 24 fps | 24 fps | RA |
| P0xS06C1Rn | BalloonFestival | 24 fps | 24 fps | RA |
| P0xS07C2Rn | WarmNight | 24 fps | 24 fps | RA |

# Results (tables)

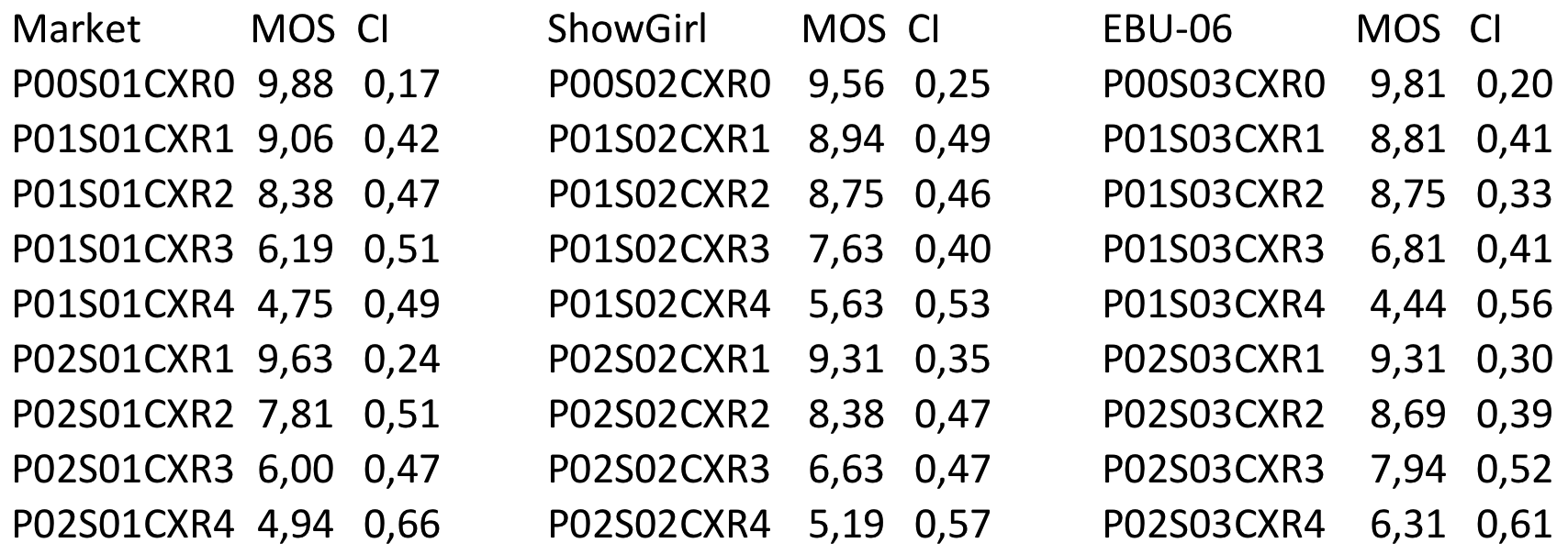


Table 3 Results for the sequences Market, ShowGirl and EBU-06

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| EBU-04 | MOS | CI |  | Garage | MOS | CI |  | Balloon | MOS | CI |
| P00S04CXR0 | 9,31 | 0,30 |  | P00S05CXR0 | 9,31 | 0,39 |  | P00S06CXR0 | 9,56 | 0,25 |
| P01S04CXR1 | 9,38 | 0,30 |  | P01S05CXR1 | 8,38 | 0,35 |  | P01S06CXR1 | 8,13 | 0,62 |
| P01S04CXR2 | 7,94 | 0,38 |  | P01S05CXR2 | 6,69 | 0,46 |  | P01S06CXR2 | 6,75 | 0,46 |
| P01S04CXR3 | 7,63 | 0,59 |  | P01S05CXR3 | 4,63 | 0,47 |  | P01S06CXR3 | 5,00 | 0,47 |
| P01S04CXR4 | 4,31 | 0,61 |  | P01S05CXR4 | 2,63 | 0,56 |  | P01S06CXR4 | 3,75 | 0,38 |
| P02S04CXR1 | 9,31 | 0,39 |  | P02S05CXR1 | 8,56 | 0,47 |  | P02S06CXR1 | 7,50 | 0,47 |
| P02S04CXR2 | 8,94 | 0,42 |  | P02S05CXR2 | 8,06 | 0,46 |  | P02S06CXR2 | 6,06 | 0,46 |
| P02S04CXR3 | 7,94 | 0,49 |  | P02S05CXR3 | 5,75 | 0,46 |  | P02S06CXR3 | 4,81 | 0,51 |
| P02S04CXR4 | 6,38 | 0,56 |  | P02S05CXR4 | 2,94 | 0,52 |  | P02S06CXR4 | 3,44 | 0,47 |

Table 4 Results for the sequences and EBU-04, Garage and Balloon

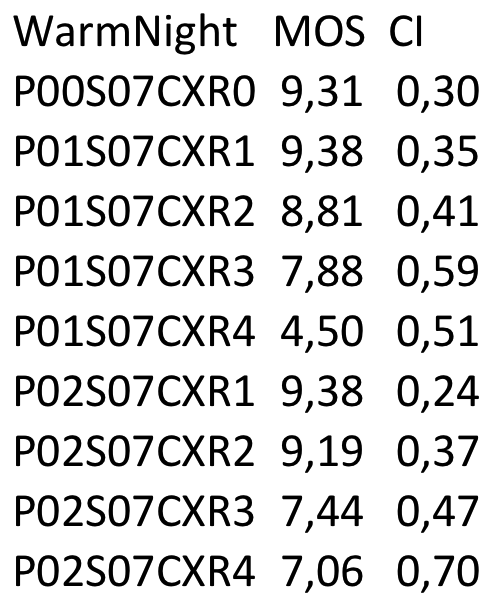


Table 5 Results for the sequences WarmNight

# Results (graphs)

The figures here below show the results of the test in form of graphs.

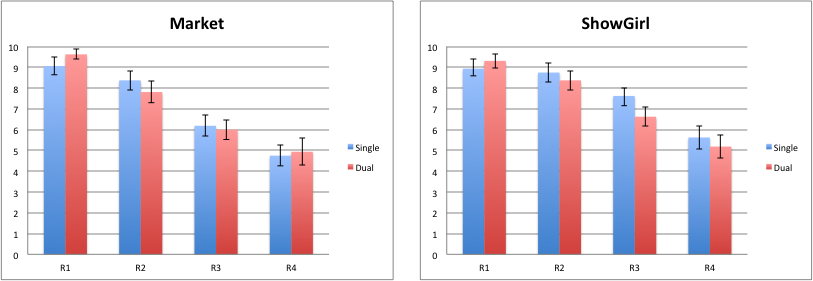


Figure 1 Results for the sequences Market and ShowGirl

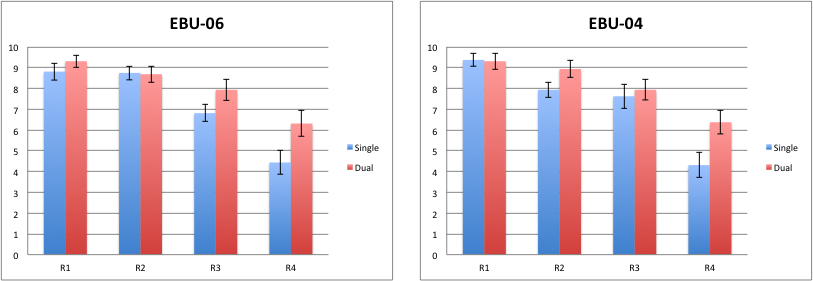


Figure 2 Results for the sequences EBU-06 and EBU-04

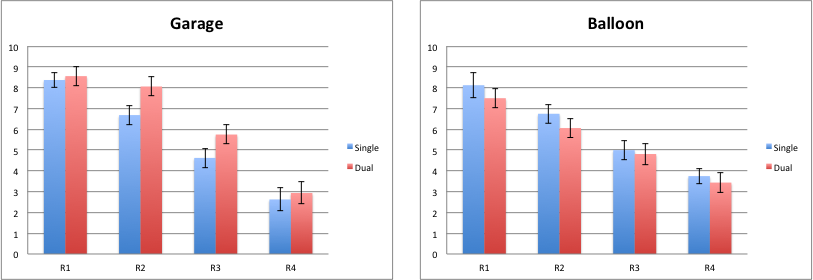


Figure 3 Results for the sequences Garage and Balloon

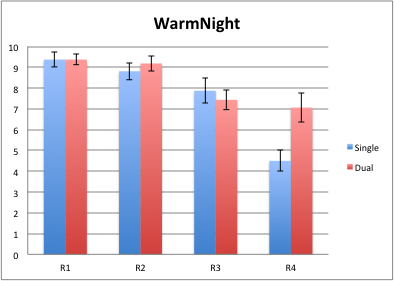


Figure 4 Results for the sequences WarmNight

## Plots of MOS vs. bitrates

In this section, the average bitrate overhead of SHVC compared to HEVC simulcast for each sequence were computed from the MOS vs. bitrate data to further quantify the bitrate savings that was achieved. The plots of MOS vs. bitrates are provided in Figure 5.

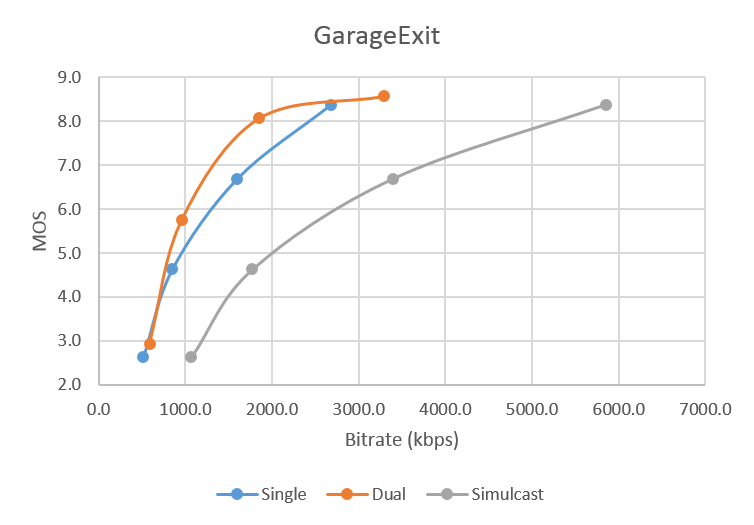
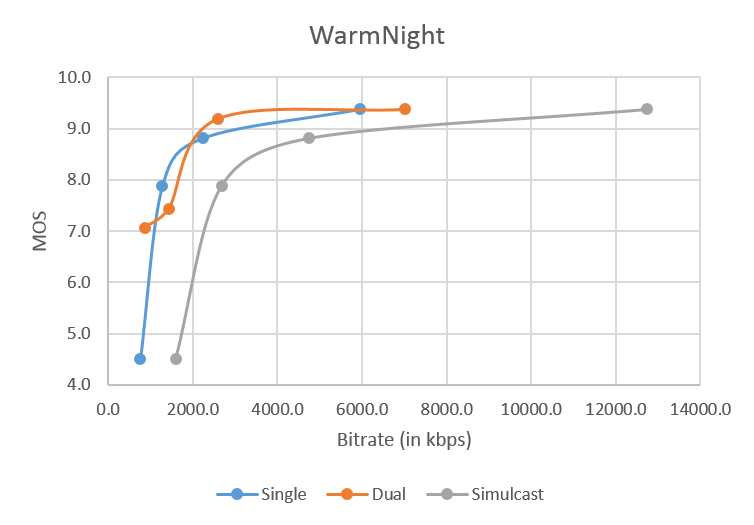
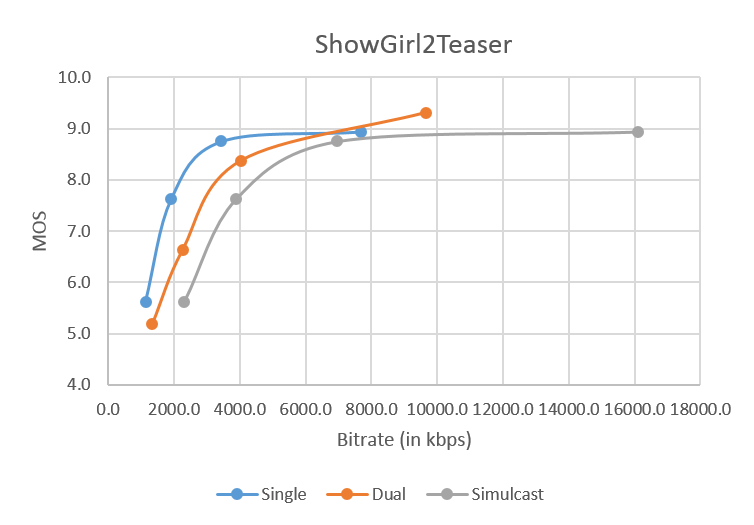
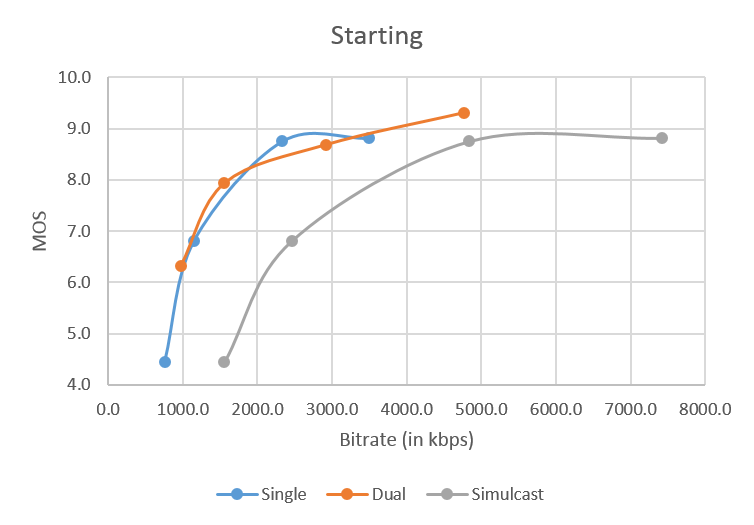
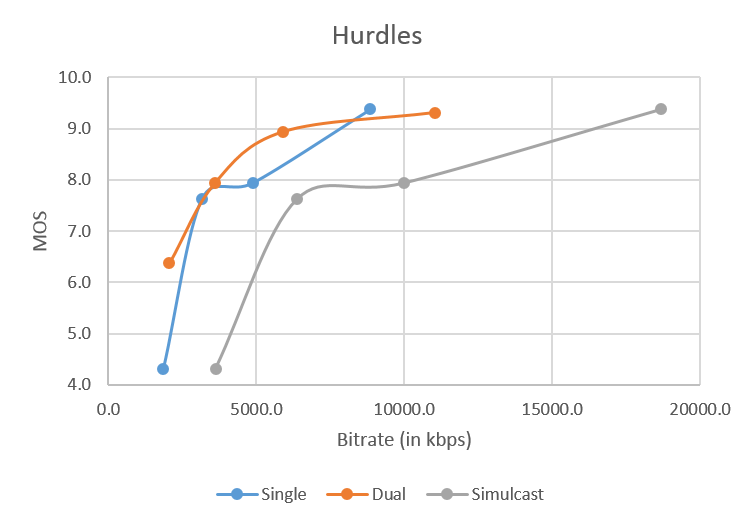
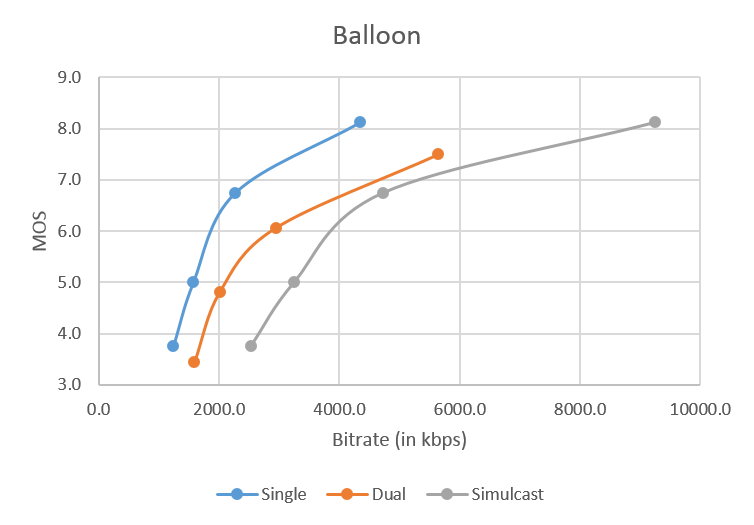
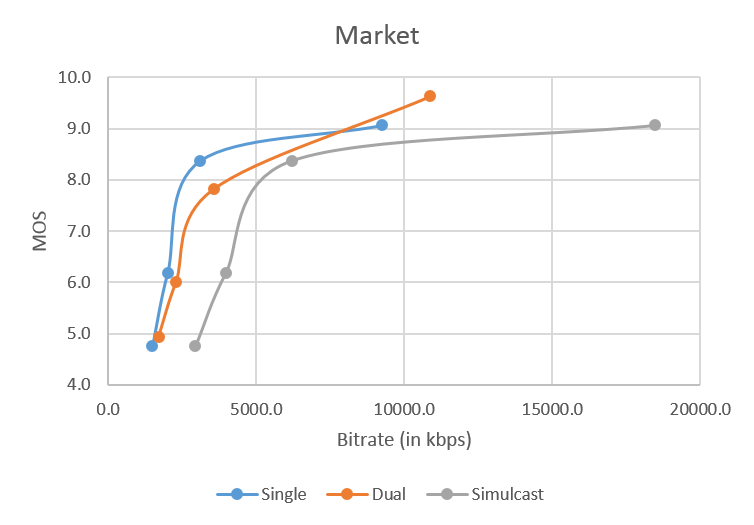


Figure 5: Average bitrate savings of SHVC vs. HEVC simulcast

All plots above show that the rate distortion (RD) curves of the SHVC test points are located to the left of the RD curves of the simulcast test points. This shows that SHVC is achieving a greater bitrate reduction over simulcast.

Table 8‑1 shows the MOS BD-rate for the sequences of each test category in this test. The BD-rate measure described in [5][6] is used with MOS scores taking the place of PSNR.

Table 8‑1 MOS BD-rate

|  |  |  |
| --- | --- | --- |
| Test sequences | BD-rate using MOS (SHVC BL+EL vs HEVC single layer) | BD-rate using MOS (SHVC BL+EL vs HEVC simulcast) |
| Market | 49.0% | -28.2% |
| BalloonFestival | 23.1% | -37.9% |
| EBU\_04\_Hurdles | 55.6% | -23.5% |
| EBU\_06\_Starting | 6.6% | -49.3% |
| ShowGirl | -16.3% | -59.0% |
| WarmNight | 3.5% | -51.2% |
| GarageExit | -16.6% | -60.5% |

**Annex A**

### *Description of testing environment and methodology*

One of the test methods described in [2] are planned to be used, applying some modification to them, in relation to the kind of display, the video recording and play-back equipment.

A.1 Test method

The test method adopted for this evaluation is DCR (Degradation Category Rating) [2].

A.1.1 Degradation Category Rating (DCR)

This test method is commonly adopted when the material to be evaluated shows a range of visual quality that well distributes across all quality scales.

This method will be used under the schema of evaluation of the quality (and not of the impairment); for this reason a quality rating scale made of 11 levels will be adopted, ranging from "0" (lowest quality) to "10" (highest quality). The test will be held in different laboratories located in countries speaking different languages: This implies that it is better not to use categorical adjectives (e.g. excellent good fair etc.) to avoid any bias due to a possible different interpretation by naive subjects speaking different languages.

The structure of the Basic Test Cell (BTC) of DCR method is made by two consecutive presentations of the video clip under test; at first the original version of the video clip is displayed, immediately afterwards the coded version of the video clip is presented; then a message displays for 5 seconds asking the viewers to vote (see )

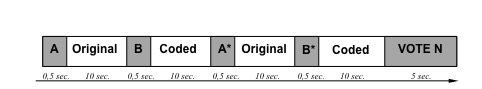


Figure 1 *- DCR (Variant II) Basic Test Cell*

The presentation of the video clips will be preceded by a mid-grey screen displaying for one second.

A.2 How to express the visual quality opinion with DCR

The viewers will be asked to express their vote putting a mark on a scoring sheet.

The scoring sheet for a DCR test is made of a section for each BTC; each section has a box wherein which the viewer shall write the score ranging from 0 to 10. By writing a score of “10”, the subject will express an opinion of “best” quality, while by writing a score of “0” the subject will express an opinion of “worst” quality.

The vote has to be written when the message "Vote N" appears on the screen. The number "N" is a numerical progressive indication on the screen aiming to help the viewing subjects to use the appropriate box of the scoring sheet.

A.4 Training and stabilization phase

The outcome of a test is highly dependent on a proper training of the test subjects.

For this purpose, each subject has to be trained by means of a short practice (training) session.

The video material used for the training session must be different from those of the test, but the impairments introduced by the coding have to be as much as possible similar to those in the test.

The stabilization phase uses the test material of a test session; three BTCs, containing one sample of best quality, one of the worst quality and one of medium quality, are duplicated at the beginning of the test session. By this way, the test subjects have an immediate impression of the quality range they are expected to evaluate during that session.

The scores of the stabilization phase are discarded. Consistency of the behaviour of the subjects will be checked inserting in the session a BTC in which original is compared to original.

A.5 The laboratory set-up

The test was done using a prototype SIM2 HDR display for which the manufacturer declared a peak luminance close to 9.000 nits in the “HDR-mode”, over the 40% of the whole screen.

A screen luminance uniformity measurement was done by means of an X-Rite luminance meter (model i1Display Pro), getting a mean luminance value around 6.400 nits, with a variation around to 2% across the 9 quadrants of the screen (up-left, up-center, up-right, center-left, center-center, center-right, down-left, down-center, down-right). The measurement was done with the display in the “non HDR” mode.

All the test signals (included the caption messages: A B A\* B\* and VOTE N) were sent to the monitor through the LogLuv DVI input. The video signals were all in AVI LogLuv file format, while the caption messages were bitmap fixed images, whose presence on the screen was set by means of a script of the MUP (Multimedia Universal Player) HDR player.

A.5.1 Viewing distance

The viewing distance was 2.5 H, where H is equal to the height of the screen. Two subjects were seated in front of the SIM2 monitor at the same time.

A.5.2 Viewing environment

The test laboratory has to be carefully protected from any external visual or audio pollution.

Internal general light has to be low (just enough to allow the viewing subjects to fill out the scoring sheets) and a uniform light has to be placed behind the monitor, in a way no direct light hits the viewing subjects seated in front of the screen; a LED lamp (AwoX SmartLight White; <http://store.awox.com/fr/p_awox_product.aspx?i=270244>) , dimmed to around 25 lumen and set for a colour temperature of 6500°, was placed behind the monitor to illuminate the wall. No other light source was on, ceiling, floor and walls of the laboratory were made of non-reflecting material.

A.6 Statistical analysis, viewers screening and presentation of the results

The data collected from the score sheets, filled out by the viewing subjects, will be stored in an Excel spread sheet.

For each coding condition the Mean Opinion Score (MOS) and associated Confidence Interval (CI) values will be given in the spread-sheets.

The MOS and CI values will be used to draw graphs. The Graphs will be drawn grouping the results for each video test sequence. No graph grouping results from different video sequences will be considered.

A screening of the viewing subjects will be done prior and after the test. Pre-screening will check for visual acuity (Snellen Chart) and colour blindness (Ishihara tables). Post screening will be done computing the correlation index of each individual viewer against the MOS values. A threshold of 0.75 will be set as lower correlation index.

From the “raw” data subject reliability should be calculated and the method used to assess subject reliability should be reported. Some criteria for subjective reliability are given in [3] and [4].

**Annex B**

### *Procedure applied to produce test bitstreams*

The following steps describe how to produce the anchor 3.2 configuration bitstreams:

1. Download HDRTools using SVN from <http://wg11.sc29.org/svn/repos/Explorations/XYZ/HDRTools/tags/0.10/>
2. Download the updated CTC package. Go to <http://wg11.sc29.org/> using a browser. Click on “MPEG-content” and navigate to [http://wg11.sc29.org/content/Explorations/HDR/CEs\_Software\_113/CE1/](http://wg11.sc29.org/content/Explorations/HDR/CEs_Software_113/CE1/updatedCTCsPackage_20160127.zip). Download updatedCTCsPackage\_20160127.zip.
3. Update HDRTools and its config files with HDRTools-0.10-src.patch and HDRTools-0.10-cfg.patch from the updated CTCs package.
4. Convert source sequences to Y’CbCr using the patched HDRTools.
5. Compile HM\_anchor3.2 from the updated CTC package and encode the sequences using config files encoder\_randomaccess\_main10\_classAGH\_vui\_lumaQP\_r1.cfg for the BT.709 sequences and encoder\_randomaccess\_main10\_classBCD\_vui\_lumaQP\_r1.cfg for the P3 sequences.

The following steps describe how to produce the SHM configuration bitstreams:

1. Download HDRTools using SVN from <http://wg11.sc29.org/svn/repos/Explorations/XYZ/HDRTools/tags/0.10/>
2. Download the SHM software patch of JCTVC-X0035 [8] and apply it to SHM-11.0.
3. Update HDRTools and its config files with HDRTools-0.10-src.patch and HDRTools-0.10-cfg.patch from the updated CTCs package
4. Convert HDR source sequences (enhancement layer) to Y’CbCr using the patched HDRTools.
5. Convert source SDR BT.709 sequences (base layer) using the patched HDRTools and first the config file HDRConvert\_Step1\_2\_YCbCr420ToRGBprime444.cfg and then the config file HDRConvert\_Step3\_RGB444toYUV420.cfg (as per JCTVC-W0046).
6. Compile the patched SHM-11.0 encode the sequences using config files encoder\_SRA\_m10\_vui.cfg and layers\_vui\_sei.cfg.

# References

1. Jonatan Samuelsson, “Conversion and Coding Practices for HDR/WCG Video, Draft 1”, JCTVC-W1017, San Diego, USA, Feb. 2016
2. International Telecommunication Union Standardization Sector; Recommendation ITU-T P.910 “Subjective video quality assessment methods for multimedia applications”
3. Pseudo Isochromatic Plates, engraved and printed by The Beck Engraving Co., Inc., Philadelphia and New York, United States.
4. KIRK (R.E.): Experimental Design – Procedures for the Behavioural Sciences, 2nd Editions, Brooks/Cole Publishing Co., California, 1982.
5. E. Francois, J. Sole, P. Yin, J. Ström, “Updated Common Test Conditions for HDR/WCG Video Coding Experiments”, JCTVC-W1020, San Diego, USA, Feb. 2016
6. ITU-T Recommendation P.910 “Subjective video quality assessment methods for multimedia applications” 04/2008.
7. Y. Ye, V. Baroncini, Y.-K. Wang, “SHVC Verification test report,” JCTVC-W1004, San Diego, USA, Feb. 2016.
8. A.K. Ramasubramonian, J. Sole, D. Rusanovskyy, D.B. Sansli, “SHM encoder improvements for HDR verification testing”, JCTVC-X0035, Geneva, CH, May-June, 2016
9. A.K. Ramasubramonian, J. Sole, Y.-K. Wang, V. Baroncini, “Draft supplemental SHVC Verification test plan,” JCTVC-X0046, Geneva, CH, May–June 2016.