|  |  |
| --- | --- |
| **Joint Collaborative Team on Video Coding (JCT-VC)**  **of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11**  16th Meeting: San José, US, 9–17 Jan. 2014 | Document: JCTVC-P1011 MPEG document: m32653 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **HEVC verification test plan** | | |
| *Status:* | Output Document from JCT-VC | | |
| *Purpose:* | Information | | |
| *Author(s) or Contact(s):* | TK Tan | Tel: Email: | +65 9799 3861 tktan@m-sph.com |
|  | Vittorio Baroncini Marta Mrak | Email: | [vittorio@fub.it](mailto:vittorio@fub.it)  Marta.Mrak@bbc.co.uk |
|  |  |
| *Source:* | AHG on verification test preparation | | |

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

# Introduction

This document contains the plan for the video verification test to be conducted to verify the coding performance of the HEVC Main and Main 10 profiles. A formal subjective evaluation will be conducted comparing the HEVC Main and Main 10 profiles to the AVC High and High 10 profiles, respectively. A range of video resolutions from 480p to 4K will be tested.

# Schedule

The following schedule is anticipated for the subjective test.

Test Sequences Identified: January 20, 2014

Encoding of sequences completed: January 27, 2014

Subjective evaluation starts: February 3, 2014

Subjective evaluation completed: March 3, 2014

Subjective evaluation results available: March 24, 2014.

# Test sites

The subjective test shall be carried out at multiple test sites. The following test sites have been identified.

* British Broadcasting Corporation (BBC) R&D (4K)
* École polytechnique fédérale de Lausanne (EPFL) (4K [max 30 fps], 1080p, 720p, WVGA(480p))
* Fondazione Ugo Bordoni (FUB) (4K [max 30 fps], 1080p, 720p, WVGA(480p))

# Test conditions

The following test conditions will be used for the HEVC verification test.

1. Number of sequences and video resolutions:
   1. 5 sequences for each resolution (480p, 720p, 1080p and 4K)
2. Bitstreams
   1. Generated with HM 12.1 for HEVC bitstreams
   2. Generated with JM 18.5 for AVC bitstreams
   3. In addition to a. and b., other HEVC and/or AVC bitstreams generated with encoders that are optimized for subjective quality may be tested if available.
3. Encoding parameters
   1. Fixed QP.
      1. 4 bit rate points per sequences covering the whole MOS range as much as possible
   2. Bit depth
      1. 8 bits for 480p, 720p and 1080p
      2. 8 and 10 bits for 4K
   3. Coding structure depending on the nature of the sequence.
      1. Random access, RA (Storage/Streaming)
         1. Intra refresh at approximately 1 second intervals.
         2. Picture reordering allowed.
      2. Low delay, LD (Video conferencing)
         1. No Intra refresh.
         2. Without picture reordering.
   4. Other settings as in the configuration files
      1. cfg/encoder\_randomaccess\_main.cfg, encoder\_randomaccess\_main10.cfg or encoder\_lowdelay\_main.cfg for HM
      2. bin/HM-like/encoder\_JM\_RA\_B\_HE.cfg or bin/HM-like/encoder\_JM\_LB\_HE.cfg configurations for JM18.5

# Test Sequences

The following test sequences are selected for the subjective test.

Table : Selected test sequences and properties

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Sequence | Source [Copyright] | Width x Height | Frame rate | Bit depth | Length (frames) | RA / LD |
| BT709Birthday | Technicolor [C3] | 3840x2160 | 50 | 10 | 500 | RA |
| Book | BBC [C4] | 3840x2160 | 50 | 10 | 500 | RA |
| manage | 4EVER [C2] | 3840x2160 | 60 | 8 | 600 | RA |
| HomelessSleeping | Kamerawerk [C8] | 3840x2160 | 60 | 8 | 600 | RA |
| traffic | Plannet, Inc [C1] | 4096x2048 | 30 | 8 | 300 | RA |
| JohnnyLobby | Vidyo [C7] | 1920x1080 | 60 | 8 | 600 | LD |
| Calendar | BBC [C4] | 1920x1080 | 50 | 8 | 500 | RA |
| SVT15 | SVT [C6] | 1920x1080 | 50 | 8 | 500 | RA |
| sedofCropped | 4EVER [C2] | 1920x1080 | 60 | 8 | 600 | RA |
| UnderBoat1 | NTIA [C5] | 1920x1080 | 30 | 8 | 300 | RA |
| ThreePeople | Vidyo [C7] | 1280x720 | 60 | 8 | 600 | LD |
| QuarterBackSneak1 | NTIA [C5] | 1280x720 | 30 | 8 | 300 | RA |
| BT709Parakeets | Technicolor [C3] | 1280x720 | 50 | 8 | 500 | RA |
| SVT01a | SVT [C6] | 1280x720 | 50 | 8 | 500 | RA |
| SVT04a | SVT [C6] | 1280x720 | 50 | 8 | 500 | RA |
| Cubicle | Vidyo [C7] | 832x480 | 30 | 8 | 300 | LD |
| Anemone | NTIA [C5] | 832x480 | 30 | 8 | 300 | RA |
| BT709BirthdayFlash | Technicolor [C3] | 832x480 | 50 | 8 | 500 | RA |
| Ducks | Plannet, Inc [C1] | 832x480 | 60 | 8 | 600 | RA |
| WheelAndCalendar | BBC [C4] | 832x480 | 50 | 8 | 500 | RA |

# Encoding Results

The following table shows the JM18.5 and HM12.1 encoding results on the sequences shown in Table 1. The QP parameters were selected such that the bitrate of the HM12.1 bitstreams are approximately half of the bitrate of the corresponding JM18.5 bitstreams. The range of the QP values was also selected so that the subjective quality of the encoded sequence span as large a range of the MOS range as possible.

Table : JM18.5 and HM11.0 encoding results

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | JM18.5 | | HM12.1 | |  |
|  |  | QPISlice | kbps (a) | QPISlice | kbps (b) | Bitrate Difference (a - b)/b |
| 4K | BT709Birthday | 24 | 14064 | 27 | 6838 | 51% |
|  |  | 30 | 7223 | 32 | 3654 | 49% |
|  |  | 35 | 4461 | 37 | 2154 | 52% |
|  |  | 40 | 2858 | 42 | 1317 | 54% |
|  | Book | 22 | 10911 | 24 | 5742 | 47% |
|  |  | 27 | 5805 | 29 | 2738 | 53% |
|  |  | 32 | 3525 | 33 | 1643 | 53% |
|  |  | 37 | 2214 | 37 | 1042 | 53% |
|  | HomelessSleeping | 23 | 38876 | 25 | 16608 | 57% |
|  |  | 26 | 12168 | 27 | 5526 | 55% |
|  |  | 31 | 5617 | 31 | 2581 | 54% |
|  |  | 37 | 3112 | 35 | 1488 | 52% |
|  | menage | 27 | 36607 | 31 | 17840 | 51% |
|  |  | 31 | 21261 | 35 | 10466 | 51% |
|  |  | 35 | 12731 | 39 | 6139 | 52% |
|  |  | 38 | 8819 | 42 | 4021 | 54% |
|  | traffic | 27 | 13309 | 31 | 6205 | 53% |
|  |  | 32 | 6583 | 36 | 3137 | 52% |
|  |  | 37 | 3618 | 40 | 1844 | 49% |
|  |  | 42 | 2090 | 44 | 1056 | 49% |
| 1080p | JohnnyLobby | 23 | 2761 | 24 | 1477 | 46% |
|  | (low delay) | 27 | 895 | 28 | 445 | 50% |
|  |  | 31 | 468 | 32 | 227 | 51% |
|  |  | 35 | 298 | 36 | 139 | 54% |
|  | Calendar | 23 | 3057 | 26 | 1407 | 54% |
|  |  | 27 | 1668 | 30 | 787 | 53% |
|  |  | 32 | 958 | 34 | 487 | 49% |
|  |  | 36 | 686 | 38 | 322 | 53% |
|  | SVT15 | 28 | 6805 | 31 | 3549 | 48% |
|  |  | 32 | 3767 | 35 | 1903 | 49% |
|  |  | 36 | 2214 | 39 | 1028 | 54% |
|  |  | 41 | 1109 | 43 | 547 | 51% |
|  | sedofCropped | 27 | 13762 | 31 | 6345 | 54% |
|  |  | 31 | 6726 | 35 | 3165 | 53% |
|  |  | 35 | 3462 | 39 | 1619 | 53% |
|  |  | 39 | 1863 | 42 | 971 | 48% |
|  | UnderBoat1 | 24 | 4407 | 27 | 1910 | 57% |
|  |  | 29 | 2026 | 31 | 990 | 51% |
|  |  | 33 | 1196 | 35 | 554 | 54% |
|  |  | 37 | 729 | 39 | 325 | 55% |
| 720p | ThreePeople | 25 | 1414 | 28 | 648 | 54% |
|  | (low delay) | 29 | 739 | 32 | 346 | 53% |
|  |  | 33 | 433 | 36 | 200 | 54% |
|  |  | 38 | 240 | 40 | 117 | 51% |
|  | BT709Parakeets | 26 | 1151 | 30 | 553 | 52% |
|  |  | 30 | 709 | 34 | 333 | 53% |
|  |  | 33 | 499 | 37 | 232 | 54% |
|  |  | 37 | 332 | 40 | 161 | 52% |
|  | QuarterBackSneak | 22 | 3844 | 25 | 1959 | 49% |
|  |  | 27 | 2039 | 30 | 1009 | 51% |
|  |  | 32 | 1145 | 35 | 541 | 53% |
|  |  | 37 | 694 | 39 | 336 | 52% |
|  | SVT01a | 27 | 1271 | 31 | 594 | 53% |
|  |  | 31 | 733 | 35 | 336 | 54% |
|  |  | 35 | 435 | 38 | 215 | 51% |
|  |  | 39 | 283 | 41 | 132 | 53% |
|  | SVT04a | 28 | 4178 | 32 | 2154 | 48% |
|  |  | 31 | 2665 | 35 | 1361 | 49% |
|  |  | 34 | 1699 | 38 | 849 | 50% |
|  |  | 37 | 1072 | 41 | 503 | 53% |
| 480p | Cubicle | 22 | 1014 | 24 | 505 | 50% |
|  | (low delay) | 25 | 502 | 27 | 264 | 48% |
|  |  | 30 | 210 | 32 | 106 | 49% |
|  |  | 35 | 105 | 37 | 49 | 53% |
|  | Anemone | 25 | 990 | 29 | 478 | 52% |
|  |  | 29 | 581 | 33 | 271 | 53% |
|  |  | 33 | 353 | 37 | 164 | 54% |
|  |  | 38 | 202 | 41 | 99 | 51% |
|  | BT709BirthdayFlash | 29 | 1515 | 33 | 774 | 49% |
|  |  | 32 | 1003 | 37 | 499 | 50% |
|  |  | 35 | 679 | 41 | 315 | 54% |
|  |  | 39 | 399 | 44 | 205 | 49% |
|  | Ducks | 27 | 2178 | 31 | 1033 | 53% |
|  |  | 32 | 1063 | 36 | 517 | 51% |
|  |  | 35 | 719 | 39 | 344 | 52% |
|  |  | 38 | 492 | 42 | 226 | 54% |
|  | WheelAndCalender | 22 | 1129 | 25 | 519 | 54% |
|  |  | 27 | 513 | 30 | 243 | 53% |
|  |  | 32 | 295 | 35 | 136 | 54% |
|  |  | 37 | 190 | 39 | 93 | 51% |

### Annex A

### *Description of testing environment and methodology*

The test procedure foreseen for the formal subjective evaluation will consider two main requirements:

* to be as much as possible reliable and effective in verifying the performance in terms of subjective quality (and therefore adhering the existing recommendations);
* to take into account the evolution of technology and laboratory set-up oriented to the adoption of FPD (Flat Panel Display) and video server as video recording and playing equipment.

Therefore, one of the test methods described in [1] 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) [1].

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 three 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.

All the video material used for these tests will consist of video clips of 10 seconds duration.

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***)



*Figure 1 - DCR BTC*

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 (see *Figure* ***2***). 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.

Scoring sheet Class C IVC english

*Figure 2 -Example of DSIS test method 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 laboratory for a subjective assessment will be set up according to [1], except for the selection of the display and the video play-out server.

For 4K video clips, high quality LCD monitors will be used with diagonal size equal to or higher than 56'' and able to accept resolutions of up to 3840x2160. Play-out of 3840x2048 video clips is done at the native resolution using the central area of the screen; the remaining part of the screen is set to a mid grey level (128 in 0-255 range)". In the case where the width of the sequence exceeds 3840, the left and right sides of the picture would be cropped and only the centre 3840 pixels are shown.

For other resolutions, High quality LCD monitors (or TV set) are used, having a diagonal size equal or higher of 40” and capable to accept resolution equal to 1920 x 1080. When using TV sets all the local colour and contrast features must be disabled (where applicable).

Play-out of 1080p, 720p and 480p video clips is done at the native resolution using the central area of the screen; the remaining part of the screen is set to a mid grey level (128 in 0-255 range).

The video play server, or the PC, used to play video has to be able to support the display of 4K, 1080p, 720p and 480p video formats, at 24, 30, 50 and 60 frames per second, without any limitation, or without introducing any additional temporal or visual degradation.

A.5.1 Viewing distance

The viewing distance varies according to the physical dimensions of the active part of the video; this will lead to a viewing distance varying from 1.5H to 4H, where H is equal to the height of the active part of the screen, depending on the size of the active part of the screen and its native resolution.

The number of subjects seating in front of the monitor is a function of the monitor size and of the selected viewing distance.

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; the light behind he monitor must be dimmed to an intensity as specified in Table 4 of Recommendation ITU-T P.911 (“Typical viewing and listening conditions as used in audio-visual quality assessment”). No other light source is admitted, and in particular any light source directed to the screen or creating reflections; ceiling, floor and walls of the laboratory have to be made of non-reflecting material (e.g. carpet or velvet) and should have a colour tuned as close as possible to mid grey.

A.6 Overall test effort and subjects’ involvement

The duration of the test will depend on the number sequences tested in each category / resolution assigned to the test laboratories; in any case each viewing session will not run for more than 20 minutes and the same viewing subject will not participated to the test run for more than six hours in total. The same subject may not be enrolled for two consecutive days. Young humans subjects, equally distributed in gender, are hired, selecting them for an age from 18 to 30 and, highly preferably among University students of scientific faculties. Viewing subjects are compensated for their participation to the testing activities (compensation may be done in money or services).

A.7 Statistical analysis 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.

Five spread-sheets will be prepared: four containing the results for 4K, 1080p, 720p and 480p (Main profile) and one for 4K (Main 10 profile).

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.

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 [2] and [3].

As an example, the reliability of a subject, could be achieved computing the correlation index between each score provided by a subject to the general MOS value assigned for that test point; in this regard a correlation index equal or superior to 0,75 (computed making the mean of all the correlation values) could be considered as valid for the acceptance of the subject

**References:**

1. International Telecommunication Union Standardization Sector; Recommendation ITU-T P.910 “Subjective video quality assessment methods for multimedia applications”
2. Pseudo Isochromatic Plates, engraved and printed by *The Beck Engraving Co., Inc*., Philadelphia and New York, United States.
3. KIRK (R.E.): Experimental Design – Procedures for the Behavioural Sciences, 2nd Editions, *Brooks/Cole Publishing Co*., California, 1982.

Annex B

### *Copyright of test sequences*

1. ***Plannet inc.***

The test sequence and all intellectual property rights therein remain the property of the owner below. This material can only be used for the purpose of developing HEVC standards. This material cannot be distributed with charge. The owner makes no warranties with respect to the material and expressly disclaims any warranties regarding its fitness for any purpose.

Owner of these sequences:

Owner: Plannet inc.

Production: Plannet inc. and IMAGICA Corp.

1. ***4EVER Consortium***

User agrees that the Sequences and all intellectual property rights therein remain the property of the 4EVER consortium members or their licensors.

These Sequences can only be used for internal research and test work dealing with Ultra High Definition, including research and test for standardization purposes. Attributing the work to 4EVER consortium will be done by attaching the following attribution notice to the Sequences:

***“Copyright © 2012-2013, all rights reserved to the 4EVER participants and their licensors. 4EVER consortium: Orange, Technicolor, ATEME, France Télévisions, INSA-IETR, Globecast, TeamCast, Telecom ParisTech, HighlandsTechnologies Solutions, www.4ever-project.com, contact: maryline.clare@orange.com. The 4EVER research Project is coordinated by Orange and has received funding from the French State (FUI/Oseo) and French local Authorities (Région Bretagne) associated to the European funds FEDER.”***

Your attribution must not be in any way that suggests that 4EVER endorses you or your use of the video sequences.

1. ***Technicolor R&D***

Technicolor R&D France SNC (“Technicolor”) is the legal owner of the video files in YUV video file format provided in the table above (“the Content”) and retains all right, title and interest (including copyright and authorship rights) in and to this Content.

Subject to compliance with the terms and conditions set forth in the present authorization of use, Technicolor hereby grants to any member of the HEVC and SHVC standardization group (“the User”), a personal, non- transferrable, non-sub-licensable, worldwide, royalty free license under Technicolor owned or controlled copyrights to display (and to copy and modify as technically necessary) the Content solely for the purpose of User’s internal processing, testing and assessment of the Content (or if relevant for the purpose of joint processing, testing and assessment of the Content with another “User”) in order to:

* evaluate the User’s contributions to the HEVC and SHVC standards (and if relevant to any multi- standard performed in JCT-VC context)
* evaluate Technicolor’s contributions to the HEVC and SHVC standards (and if relevant to any multi- standard performed in JCT-VC context)
* evaluate other third party HEVC and SHVC standards contributors’ contributions to HEVC and SHVC standards (and if relevant to any multi-standard performed in JCT-VC context)

1. ***British Broadcasting Corporation (BBC)***

The video sequences provided above and all intellectual property rights therein remain the property of the BBC. The BBC is making available the video sequences for use under the ***Creative Commons Attribution-NonCommercial 3.0 licence***.

You are free to use, share (to copy, distribute and transmit) or remix (to adapt) the BBC video sequences, provided that:

* **No-commercial**- you may not use these video sequences for commercial purposes; and
* **Attribution**- you attribute the work to the BBC by indicating that the video sequences (or elements thereof) were produced by the BBC. Your attribution must not be in any way that suggests that the BBC endorses you or your use of the video sequences.

1. ***The National Telecommunications and Information Administration (NTIA)***

<http://www.its.bldrdoc.gov/resources/video-quality-research/video-footage.aspx>

Standards committees can use CDVL R&D content within subjective tests to validate objective video quality models (e.g., ATIS, VQEG, ITU).

1. ***Sveriges Television AB (SVT), Sweden***

Individuals and organizations extracting sequences from this archive agree that the sequences and all intellectual property rights therein remain the property of Sveriges Television AB (SVT), Sweden. These sequences may only be used for the purpose of developing, testing and presenting technology standards. SVT makes no warranties with respect to the materials and expressly disclaim any warranties regarding their fitness for any purpose.

1. ***Vidyo, Inc.***

Vidyo donates the sequences to the public domain (JCTVC-P0042)

1. ***Kamerawerk***

The video sequences provided above and all intellectual property rights therein remain the property of the Kamerawerk. The Kamerawerk is making available the video sequences for use under the ***Creative Commons Attribution-NonCommercial 3.0 licence***.