 ISO/IEC JTC 1/SC 29/WG 04 N0126

**ISO/IEC JTC 1/SC 29/WG 04  
MPEG Video Coding   
Convenorship: CN**

**Document type:** Output Document

**Title:** Common Test Conditions for Essential Video Coding

**Status:** Approved

**Date of document:** 2021-07-16

**Source:** ISO/IEC JTC 1/SC 29/WG 04

# Expected action: None

# Action due date: None

**No. of pages:** 8 (without cover page)

**Email of Convenor:** yul@zju.edu.cn

**Committee URL:** <https://isotc.iso.org/livelink/livelink/open/jtc1sc29wg4>

**INTERNATIONAL ORGANIZATION FOR STANDARDIZATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC 1/SC 29/WG 04 MPEG VIDEO CODING**

**ISO/IEC JTC 1/SC 29/WG 04 N** **0126**

**July 2021, Online**

|  |  |
| --- | --- |
| **Title** | **Common Test Conditions for Essential Video Coding** |
| **Source** | **WG 04, MPEG Video Coding** |
| **Status** | **Approved** |
| **Serial Number** | **20739** |

**Abstract**

This document provides common test conditions (CTC) of Essential Video Codec (EVC) and software reference configurations associated with the correspondent test model (ETM).

1. **Introduction**

This document on Common Test Conditions (CTC) defines test categories and encoder configuration utilized for testing of the ETM reference software of version 7.3.2 and higher. These test conditions are expected to be utilized and results reported for these test conditions for most of experiments.

Test conditions for 3 categories are defined:

1. SDR category

Coding of video content in Standard Dynamic Range and Standard Color Gamut representation, e.g. in BT.709

1. HDR/WCG category

Coding of video content in High Dynamic Range and Wide Color Gamut representation, e.g. in BT.2100.

1. SCC category

Coding of screen content, e.g. computer generated material.

Two encoder configurations are specified, reflecting random-access and low delay settings:

* Random access, 10 bit
* Low delay, 10 bit

ETM reference software of version 7.3.2 and higher are expected to be used for most experiments. Tests targeting baseline profile of MPEG 5/EVC are required to reported results in SDR category. Tests targeting main profile of MPEG 5/EVC are required to report results in SDR category. In addition, test results for HDR and SCC categories can be optionally reported. There are category-specific tools, i.e. DRA for HDR/WCG testing and IBC for SCC testing, that should be enabled in case of testing HDR/WCG and SCC categories.

The following sections define test sequences, quantization parameter values and encoder configuration files to be used.

1. **Test sequences**

Table 1 defines a set of test sequences to be used for coding of SDR material.

All frames (as defined by frame count in the table) shall be encoded for all sequences and test cases described below. Letters “M” and “O” in the Table 1 indicate mandatory and optional, respectively.

Table 1. Test sequences for SDR category

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Class | Sequence name | Frame count | Frame rate | Bit depth | Random access | Low-delay |
| A | Tango2 | 294 | 60 | 10 | M | O |
| A | FoodMarket4 | 300 | 60 | 10 | M | O |
| A | CatRobot1 | 300 | 60 | 10 | M | O |
| A | DaylightRoad2 | 300 | 60 | 10 | M | O |
| A | ParkRunning3 | 300 | 50 | 10 | M | O |
| B | MarketPlace | 600 | 60 | 10 | M | M |
| B | RitualDance | 600 | 60 | 10 | M | M |
| B | Cactus | 500 | 50 | 8 | M | M |
| B | BasketballDrive | 500 | 50 | 8 | M | M |
| B | BQTerrace | 600 | 60 | 8 | M | M |
| E | FourPeople | 600 | 60 | 8 | O | M |
| E | Johnny | 600 | 60 | 8 | O | M |
| E | KristenAndSara | 600 | 60 | 8 | O | M |

Table 2 defines a set of test sequences to be used for coding of SCC material.

Table 2. Test sequences for SCC category

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Class** | **Sequence name** | **Frame count** | **Frame rate** | **Bit depth** |
| S | FlyingGraphics | 300 | 60 | 8 |
| S | Desktop | 600 | 60 | 8 |
| S | Console | 600 | 60 | 8 |
| S | ChineseEditing | 600 | 60 | 8 |

Table 3 defines a set of test sequences to be used for coding of HDR/WCG test material.

Table 3. Test sequences HDR/WCG category

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Class** | **Sequence name** | **Frame count** | **Frame rate** | **Bit depth** | **Transfer Function** |
| H1 | BalloonFestival | 240 | 24 | 10 | PQ |
| H1 | Cosmos1\_TreeTrunk | 240 | 24 | 10 | PQ |
| H1 | EBU\_Hurdles | 500 | 50 | 10 | PQ |
| H1 | EBU\_Starting | 500 | 50 | 10 | PQ |
| H1 | Market | 400 | 50 | 10 | PQ |
| H1 | ShowGirl | 339 | 25 | 10 | PQ |
| H1 | SunRise | 200 | 25 | 10 | PQ |
| H2 | DayStreet | 300 | 60 | 10 | HLG |
| H2 | FlyingBirds2 | 300 | 60 | 10 | HLG |
| H2 | PeopleInShoppingCenter | 300 | 60 | 10 | HLG |
| H2 | SunsetBeach2 | 300 | 60 | 10 | HLG |

Table 4 defines a checksum information for the test sequences in all categories.

Table 4. MD5 checksum of SDR test sequences

|  |  |  |
| --- | --- | --- |
| **Class** | **Sequence name** | **MD5 checksum** |
| A | Tango2 | 0471a59c423b7059c5c6c8b395e864a9 |
| A | FoodMarket4 | a378b34190f54f688d048a9a8b46a8ac |
| A | CatRobot1 | 03a89792693fd9ecfd72ef2590025e97 |
| A | DaylightRoad2 | bf1d22643afb41b1521749d2749fb5f0 |
| A | ParkRunning3 | e7a1d1ebff269767ec4bffd2998d1154 |
| B | MarketPlace | dc668e7f28541e4370bdbdd078e61bba |
| B | RitualDance | a3cb399a7b92eb9c5ee0db340abc43e4 |
| B | Cactus | 3fddb71486f209f1eb8020a0880ddf82 |
| B | BasketballDrive | d38951ad478b34cf988d55f9f1bf60ee |
| B | BQTerrace | efde9ce4197dd0b3e777ad32b24959cc |
| E | FourPeople | 4ce5d72311b32acce62614f63225fba5 |
| E | Johnny | be83259b3ccdada2213fbd8dea20bf6e |
| E | KristenAndSara | aa3931974e34deba15a1018ba3bf5e0c |
| S | FlyingGraphics | 02a9d384fb4a0e0480dd23bebfe0460d |
| S | Desktop | f3c504cde8d4d6d554cc2d29b8dc2e77 |
| S | Console | e87d0fd0a88895844dfeb2e04a987fc2 |
| S | ChineseEditing | f8053fcf3a8beb623e76dd81ad4b1491 |
| H1 | BalloonFestival | a7831be36b65645ad1963bd367370141 |
| H1 | Cosmos1\_TreeTrunk | da4a2488c249720da0535f01c3693efa |
| H1 | EBU\_Hurdles | bc3cba849d6f4ee74d39056600722aa5 |
| H1 | EBU\_Starting | 1cbc416696cb0dfcf4da9886eeb6a4a2 |
| H1 | Market | c97abe47455fd12f6d6436cecfad7c7d |
| H1 | ShowGirl | 44f1974d68f7799c71eea29fb72b245b |
| H1 | Sunrise | 420202b3390998c9148bf5958501186d |
| H2 | DayStreet | 2bd7bbcfb80ab329118a0a82101e2715 |
| H2 | FlyingBirds2 | df6aa8af4575893c9d3b0a555d8f47b4 |
| H2 | PeopleInShoppingCenter | 40395778572368438c0202fe3c017fa9 |
| H2 | SunsetBeach2 | 9db879b083ad2b5aab31633ba46fc590 |

Notes:

* The MD5 checksums for FoodMarket4, CatRobot1, DaylightRoad2 and ParkRunning3 are for full length sequences. In these CTC only the first 300 frames are used.
* The sequences DayStreet, PeopleInShoppingCenter and SunsetBeach2 are distributed as 600 frame files, even though only 300 frames are required by this CTC and specified in Table 3. The MD5Sum values reported here correspond to the 600 frame files.

Test sequences of class A, H1 and H2 are available on <ftp://ftp.ient.rwth-aachen.de> and <ftp://ftp.hhi.fraunhofer.de>

Test sequences of classes B, E and S are available on <ftp://ftp.tnt.uni-hannover.de/testsequences/>.

Test sequence are only available to qualified MPEG participants. Qualified participants may contact the MPEG video chair (Prof. Lu Yu [yul@zju.edu.cn](mailto:yul@zju.edu.cn)) for login information.

1. **Quantization parameter values**

For each video sequence four quantization parameter values are to be used: 22, 27, 32 and 37. These values define the target QP values for a sequence. Adaptation of QP to lambda and refinement for frames of the GOP in the configuration files may change the used QP value on per-frame basis.

1. **Configuration files**

The following section defines encoder configuration files to be used for each test case. Parameters to be changed for each test point are:

* InputFile to reflect the location of the source video sequence on the test system;
* FrameRate to reflect the frame rate of a given sequence;
* SourceWidth to reflect the width of the source video sequence;
* SourceHeight to reflect the height of the source video sequence;
* FramesToBeEncoded to reflect the frame count of a given sequence;
* IntraPeriod to reflect the intra refresh period in the random access test cases. The intra refresh period is dependent on the frame rate of the source: a value 48 for 50fps; 64 for 60fps;
* QP to reflect the quantization parameter values defined in section 3;
* InputBitDepth to reflect the bit depth of a given sequence;

The following configuration files are provided in the cfg/ folder of the common software package.

* Cfg files for ETM in main configuration:
  + “Random access” (RA): encoder\_randomaccess.cfg
  + “Low-delay B” (LB): encoder\_lowdelay.cfg
  + For SCC test sequences a specific configuration file (per-class/classS.cfg) shall be used in order to activate intra block copy (IBC) coding tool
  + For HDR/WCG test sequences a specific configuration files (per-class/classH1.cfg) shall be used in order to activate Dynamic Range Adjustment (DRA) coding tool
* Cfg files for ETM in baseline configuration:
  + “Random access” (RA): encoder\_randomaccess\_baseline.cfg
  + “Low-delay B” (LB): encoder\_lowdelay\_baseline.cfg

1. **Anchors**

The current ETM test model in baseline configuration and JM19.0 are provided as a primary and a secondary anchors correspondently for the tests targeting to baseline configuration.

The current ETM test model in main configuration and HM16.22 are provided as a primary and a secondary anchors correspondently for the tests targeting to main configuration.

1. **Reported Metrics**

The attached Excel sheets contain a reporting template in which bitrate, PSNR, coding time and BD-rate results are reported for the tested configuration against the anchors. For HDR/WCG category, wPSNR, DE100 and L100 HDR metrics are to be reported.

For the purpose of reporting encoding and decoding running times, the anchor and test should be simulated on the same platform, i.e. the same CPU cores, to have reliable time comparison. The Excel summary sheet contains fields to report attributes of the testing environment, including CPU type, compiler, and decoder configuration. Filling in these fields when making a report is highly encouraged.

In the attached Excel sheets, BD-rate is calculated using both piece-wise cubic and cubic interpolations, and the piece-wise cubic interpolation is used in the results summary.

* 1. **PSNR**

For 10-bit video, PSNR is calculated as , 8-bit content is converted to 10-bit input in the encoder by shifting 2 bits to the left and 10-bit PSNR calculation is used to report testing results. This behaviour is built into the reference encoder and no external conversion program is required.

* 1. **wPSNR**

The wPSNR, or weighted PSNR, metric is calculated from the weighted mean squared error of the pixel values. An implementation of the metric is provided in the ETM software, and that implementation shall be used for calculation of the metric using the weighting functions provided below. It is noted that the HDRMetrics tool is capable of computing the weighted PSNR metric. If the HDRMetric package is used, it shall be configured to provide the same result as the implementation of the metric provided in the ETM software.

The wPSNR, or weighted PSNR, metric is calculated from the weighted mean squared error of the pixel values. wPSNR metric is calculated as:

,

where X is the maximum pixel value for the specific bit depth and *wMSE* is given as

,

where is a weight that is a function of the luma value corresponding to the pixel , *xorig,i* is the original value at location *i*, and *xdec,i* is the reconstructed value at location *i*.

The calculation of the weight is computed as:

In all cases, the metric is calculated individually for a single luma or chroma channel and then used to compute a Bjøntegaard Delta-Rate and Delta-PSNR rate.

* 1. **deltaE100-based metric**

An implementation of the deltaE100 metric is provided in the ETM software, and that implementation shall be used for calculation of the metric using the weighting functions provided below. It is noted that the HDRMetrics tool is capable of computing the deltaE100 metric. If the HDRMetric package is used, it shall be configured to provide the same result as the implementation of the metric provided in the ETM software. Details below are provided for information.

First the orignal and test material are converted to linear-light 4:4:4 RGB EXR. For example, if the material is in the YCbCr BT.2100 4:2:0 PQ 10 bit format, it is converted to a 4:4:4 RGB BT.2100 OpenEXR. Subsequently, the following steps should be applied for each (R,G,B) sample, within the content to be compared, i.e. original source (content 1) and test material (content 2) :

* Convert the content to the XYZ colour space
* Convert the content from the XYZ to Lab space using the following equations. Computations are performed using double floating point precision.

with convToLab(x) is defined as

* The deltaE100 distance DE between two samples (L1,a1,b1) and (L2,a2,b2) is then computed as follows:

with

After this process the deltaE100 values for each frame are averaged within the “Distortion” specified window. Finally, a PSNR based value is derived as:

where PeakValue is set to 10,000 for sequences in the PQ format and 1,000 for sequences in the HLG format.

* 1. **PSNR-L100**

An implementation of the PSNR-L100 metric is provided in the ETM software, and that implementation shall be used for calculation of the metric using the weighting functions provided below. It is noted that the HDRMetrics tool is capable of computing the PSNR-L100 metric. If the HDRMetric package is used, it shall be configured to provide the same result as the implementation of the metric provided in the ETM software. Details below are provided for information.

PSNR-L100 represents the distortion in the lightness domain of the CIELab colour space. The derivation of Lab values from the linear representation of the signal is similar as given in the description of deltaE100. The mean absolute error (MAE) in the L domain is used to compute the PSNR-L100 as follows:

where PeakValue is set to 10,000 for sequences in the PQ format and 1,000 for sequences in the HLG format.[[1]](#footnote-1)

1. **Parallel encoding/decoding**

Parallel encoding/decoding may be applied for RA configurations. If parallel encoding and decoding is used it should be mentioned and the method should be applied for both versions, the reference and the version under test.

1. As defined above, reporting the deltaE100, PSNR-L100 and wPSNR metrics is not required for HLG content. This example is provided solely for completeness. [↑](#footnote-ref-1)