

EBU

OPERATING EUROVISION AND EURORADIO

TR 038

SUBJECTIVE EVALUATION OF HYBRID LOG GAMMA (HLG) FOR HDR AND SDR DISTRIBUTION

EBU TECHNICAL REPORT

Geneva
March 2017



Executive Summary

In July 2016, the ITU standardised two system transfer characteristics for High Dynamic Range (HDR) in Television in the ITU R BT.2100 Recommendation. Both systems called the Perceptual Quantizer (PQ) and the Hybrid Log-Gamma (HLG) can be used for HDR content production and distribution.

The HLG system offers a degree of compatibility with legacy displays limited to Standard Dynamic Range (SDR) that are able to interpret the ITU R BT.2020 colour space signalling (UHD 1 Phase 1 displays).



Four independent test labs (the IRT, Orange, the RAI and the EBU) teamed up to perform subjective tests with the following goals:

1. identify if HDR visual quality of a given sequence was subjectively equivalent in PQ and in HLG in the compressed domain;
2. identify if the visual quality of an HLG bit-stream presented on an SDR display was equivalent to the visual quality of a sequence produced by a legacy SDR TV-production

The tests have been designed with the intention to simulate a short term live broadcast use case i.e. relying on a 1000 cd/m² workflow.

The first goal was assessed using two different test methodologies: DSCQS and SAMVIQ, both standardized by ITU R, with some adaptations to HDR. Results are in agreement with both methods and show that HLG-HDR provides quality results at least as good as PQ. An HDR TV service is thus expected to provide the same quality whichever HDR system is used for distribution.

Furthermore, concerning the PQ to HLG conversion, it can be noticed that this process does not impact the perceived picture quality of the source content. Original contents produced in PQ can therefore be broadcast in HLG without any conversion penalty (in HDR).

The second goal was assessed using the SAMVIQ methodology only. Results show that HLG-SDR and “SDR manual grade” provide **equivalent** quality results. Although differences between the two versions are noticeable, the user perception of the quality is the same in both cases.

Contents

Executive Summary	3
1. Introduction	7
2. Background	7
3. Main test objectives	8
4. Test Setup	9
4.1 Test material	9
4.2 Test preparation for Test 1 (HDR) and Test 2 (SDR)	9
4.2.1 Generation of the HDR reference	9
4.2.2 Conversion from PQ to HLG	10
4.2.3 SDR content grading	10
4.3 Sequences encoding	10
4.4 Viewing conditions and viewers' selection	11
4.5 Test conditions	11
4.6 Test methods to evaluate subjective picture quality	12
4.6.1 DSCQS (Test 1 only)	12
4.6.2 SAMVIQ (Test 1 and Test 2)	14
5. Test results	16
5.1 DSCQS	16
5.2 SAMVIQ	17
6. Main Conclusions	19
Annex A: Instructions to participants at the tests	21
A.1 Instruction for viewers (SAMVIQ Test 1)	21
A.2 Instruction for viewers (SAMVIQ Test 2)	22
A.3 Instruction for viewers (DSCQS Test 1)	23
Annex B: Detailed Test Results	25
B.1 Test 1 (HDR)	25
B.1.1 IRT detailed DSCQS test results for HDR	25
B.1.2 Orange/RAI detailed SAMVIQ test results for HDR	28
B.2 Test 2 (SDR)	32
B.2.1 Orange/RAI detailed test results SAMVIQ for HDR	32

Subjective evaluation of Hybrid Log Gamma (HLG) for HDR and SDR distribution

<i>EBU Committee</i>	<i>First Issued</i>	<i>Revised</i>	<i>Re-issued</i>
TC	2017		

Keywords: High Dynamic Range, HDR, Hybrid Log Gamma, HLG, SDR compatibility, Distribution

1. Introduction

High Dynamic Range (HDR) is considered a key feature for the success of new UHDTV services. The recent ITU-R BT.2100 Recommendation defines two possible system transfer characteristics, the Perceptual Quantizer (PQ) and the Hybrid Log Gamma (HLG).

Four independent labs, representative of the European television broadcasters and service providers, organized a campaign of subjective assessments on the perceived quality of UHD television including HDR contents encoded in HEVC as distribution format at various bit rates, as displayed by both HDR and SDR (Standard Dynamic Range) panels. This document describes the context of the tests conducted, the methodologies used, the results obtained and the conclusions drawn.

2. Background

In 2016, in the context of the global standardization efforts around HDR, various candidate technologies were considered for the distribution of HDR services, some of them claiming to be backward compatible with legacy SDR services.

Before selecting any solution, it was important to perform tests on the visual quality and the efficiency of any candidate solution.

In the absence of standardized methodology for HDR quality assessment, four independent test labs (EBU, IRT, Orange and RAI), each with a strong past experience in conducting subjective tests, joined forces to define a test plan for the evaluation of HDR technologies in distribution.

The test plan was defined jointly with broadcasters, who gave their requirements while considering short term deployments. The test plan for HDR and SDR subjective evaluation thus takes into account the broadcasters' and service providers' priorities (defining realistic test conditions, e.g. a 1000 cd/m² workflow, selection of reference monitor, choice of target bitrates...).

Given the fact that PQ-based HDR10¹ was first supported by HDR TV sets and also adopted by the Blu-ray Disc Association (BDA) as being mandatory to support HDR format, PQ was considered as the reference for HDR distribution.

Among the backward compatible HDR solutions, HLG10², proposed by BBC, was the first candidate under test.

Consequently, subjective tests were conducted to compare the performance of PQ with HLG as well as the performance of the backward compatible version with a legacy SDR production (SDR native).

¹ HDR10: specifies the use of the PQ EOTF (SMPTE ST.2084) with 10 bit quantization, colour space: ITU-R BT.2020 and MaxCLL/MaxFALL static metadata (SMPTE ST.2086). It will be referred as PQ10 when the static metadata are not used.

² HLG10: specifies the use of the Hybrid Log Gamma (OETF), with 10-bit quantization and ITU-R BT.2020 colour space.

3. Main test objectives

The four test labs conducted a subjective test campaign so as to compare PQ and HLG in a first test and to compare SDR native to the HLG backward compatible signal in a second test.

The issues addressed by these tests are the following:

- Test 1: compare the performance (with and without HEVC compression) of HLG with PQ in terms of perceived video quality on an HDR reference monitor.
- Test 2: compare the performance (with and without HEVC compression) of HLG with legacy SDR production in terms of perceived video quality on a SDR reference monitor.

Figure 1 depicts the live production workflows scenarios under consideration.

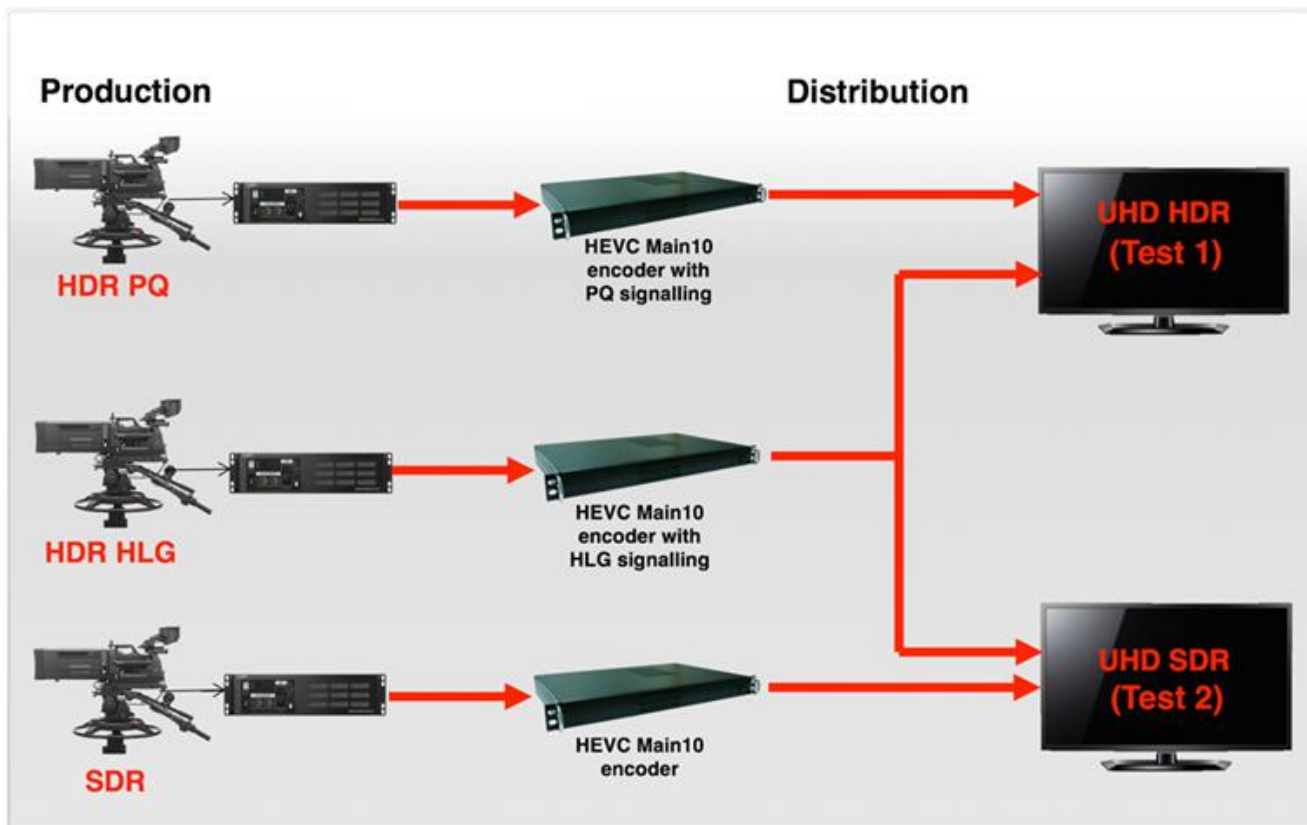


Figure 1: Live production workflows

4. Test Setup

4.1 Test material

A set of 6 test sequences representing a large range of broadcast content genres such as sport, movie and documentary was selected. The choice of sequences was made so as to ensure a large range of luminance and contrast levels (Figure 2) sequences were provided by Orange (*Movie*; courtesy of the 4Ever project, *Tennis*; courtesy of France Télévisions, and *Sailing*), the RAI (*Cycling* and *Football*) and the EBU (*Fireworks*).

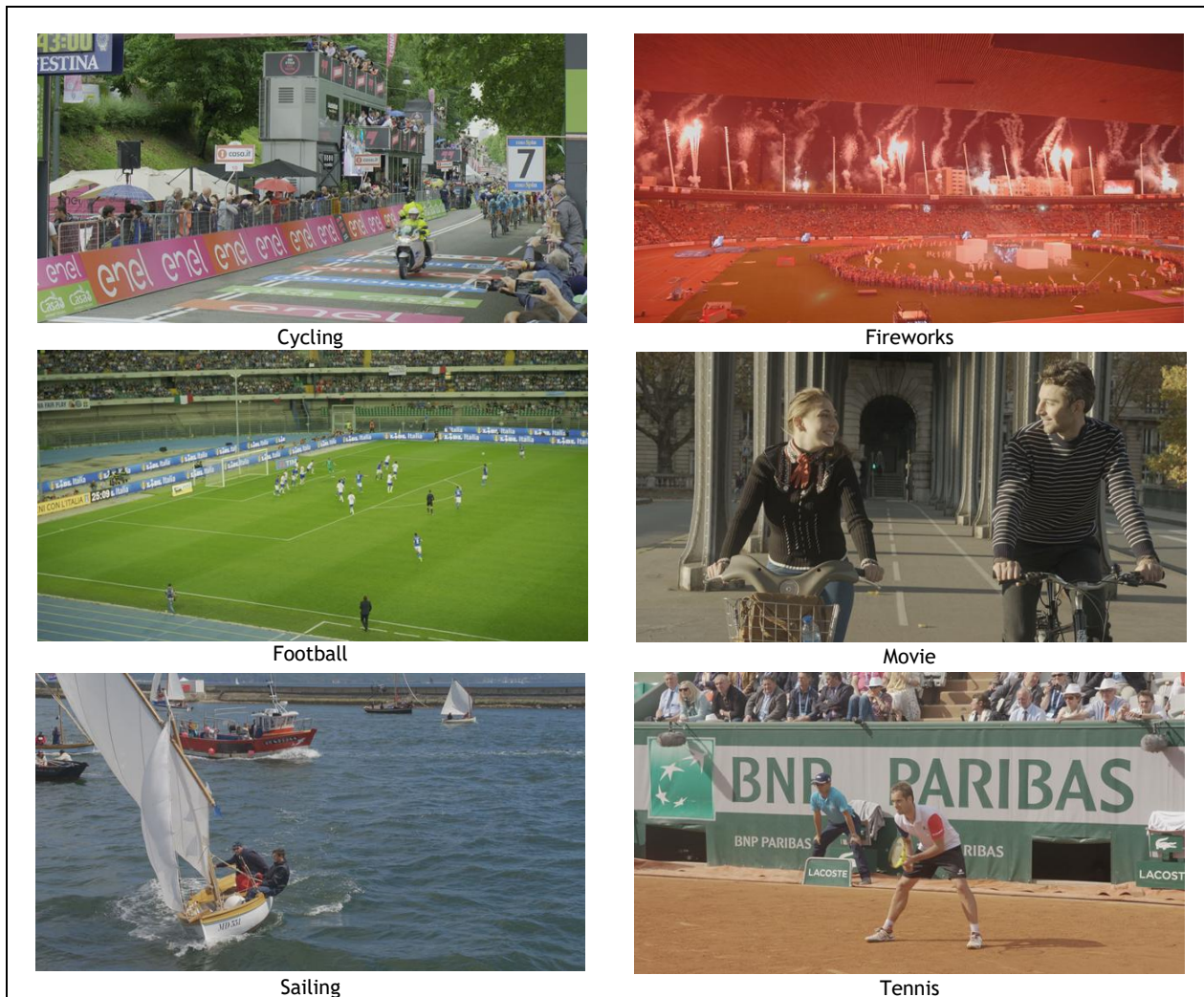


Figure 2: Source sequences

4.2 Test preparation for Test 1 (HDR) and Test 2 (SDR)

4.2.1 Generation of the HDR reference

Each source test content (shot in linear RAW) was first graded in PQ format using a DaVinci Resolve software (release 12.5) and a Sony BVM-X300 monitor for visual control. The PQ grading was done with a 1000 cd/m² peak luminance and BT.2020 colour space.

This PQ version was used as the HDR explicit and hidden reference in Test 1. The grading of the sequences was carried out by video engineers using the reference viewing environment defined for critical viewing of HDR programme material (described in Recommendation ITU-R BT.2100).

4.2.2 Conversion from PQ to HLG

Each PQ HDR reference sequence was converted into HLG following the procedures defined in BT.2100³ annex 2. Figure 3 illustrates the conversion process used. The resulting HLG uncompressed version of the sequence was used in Test 1 so as to assess the perceptual impacts of the conversion from PQ to HLG.

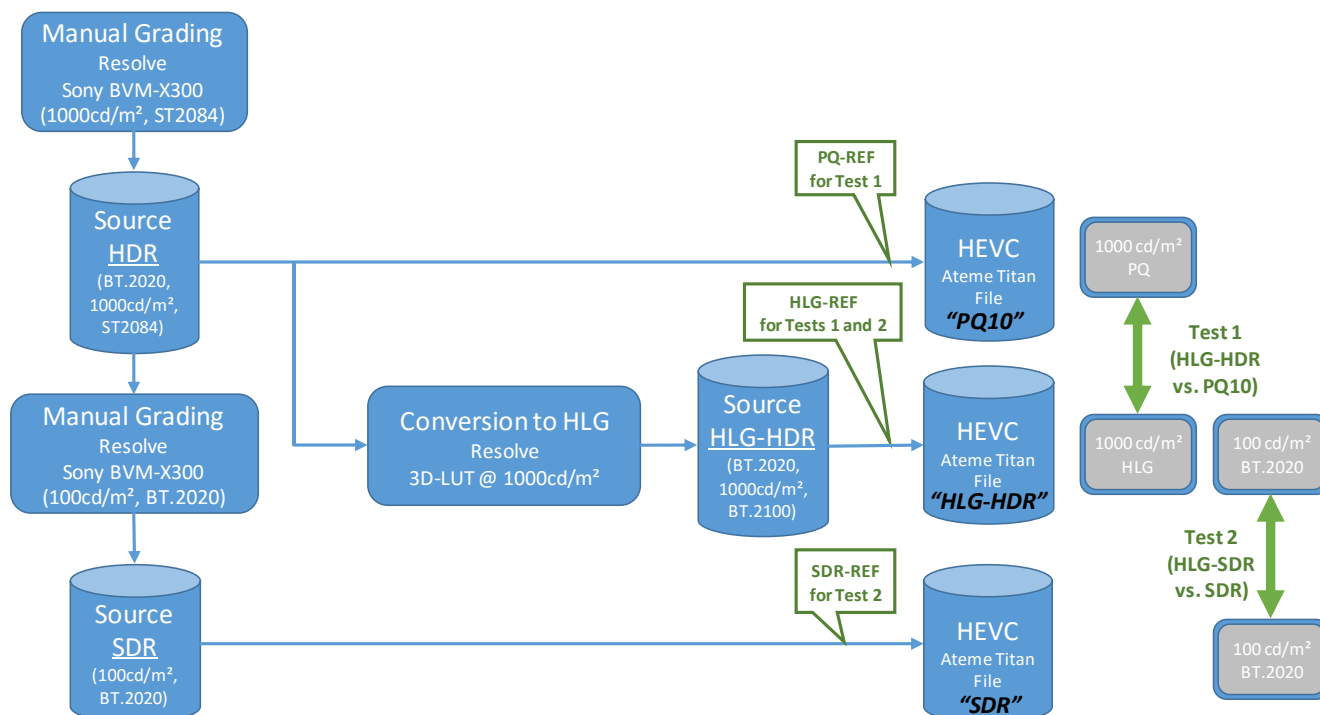


Figure 3: Production process of test material for Test 1 and Test 2

4.2.3 SDR content grading

Starting from the uncompressed HDR PQ version of the sequence, a manual SDR grading was also done on DaVinci Resolve, still using a Sony BVM-X300 monitor adjusted with a BT.2020 transfer function but with a peak luminance of 100 cd/m². In order to be representative of current SDR live TV production, the following rules were applied when necessary:

- Reduction of the dynamic in black levels
- Clipping of high level of luminance (white levels)
- Optimization of contrast on regions of interest

For the grading of the SDR sequences, the reference viewing environment was the same as for HDR (see § 4.2.1).

4.3 Sequences encoding

The graded uncompressed sequences (PQ, HLG and SDR native) were encoded using HEVC video compression. The encoding was performed using the ATEME TITAN FILE encoder (version 3.7.7 of KFE) with the following features:

- 3840 x 2160p/50, 10-bit, BT.2020 colour space
- Four different bitrates (2.5, 5, 10 and 20 Mbit/s)
- HEVC Main10 profile, Random Access Point period of max. 1 second
- SEI/VUI Messages used within the bit-stream:

³ This refers the ITU-R BT.2100 version of July 2016.

- HLG: Primaries BT.2020 and HLG Transfer Function
- PQ: Primaries BT.2020, Transfer Function SMPTE ST2084, MaxCLL: set to 1000, MaxFALL: set to 100 (MaxFALL and MaxCLL are not interpreted in the following due to uncompressed playback of the sequences during the tests).
- SDR: Primaries BT.2020, Transfer Function BT.2020

The version of the encoder that was used had an enhanced HDR psychovisual rate control dedicated to each transfer function for luminance and it also supported inclusion of the appropriate HDR metadata and signalling when applicable. This approach is typical and adequate for an advanced state-of-the-art encoder. The extent to which the psychovisual rate control adapts to a transfer function is dependent on the implementation and thus may lead to enhanced test results in the future.

4.4 Viewing conditions and viewers' selection

Both test methods (SAMVIQ and DSCQS, see § 4.6) used the same viewing environment for both SDR and HDR subjective tests.

The viewing environment for SDR subjective assessment has been defined considering the Rec. ITU-R BT.500-13. For HDR subjective test with non-expert viewers (Test 1), there is no viewing conditions standardized at the moment. Following Rec. ITU-R BT.500-13 rules, the background luminance should have been fixed to at least 100 cd/m², and assuming a 1000 cd/m² peak luminance at the display side. However, considering the reference viewing environment for critical viewing of HDR programme material as defined in Rec. ITU-R BT.2100, it was decided to keep the background luminance as the same level as for the SDR subjective test (10 cd/m²) and this value is close to the recommended one for HDR television production.

A summary of the main parameters applied:

- Viewing distance: 1.5H
- Background luminance: 10 cd/m²
- Display: Sony 30" OLED BVM-X300 used both for SDR and HDR assessment with appropriate settings.

Before each subjective test, the non-expert viewers were screened for normal visual acuity and for normal colour vision.

4.5 Test conditions

For Test 1 on HDR content (SAMVIQ), viewers had to assess the following test conditions for each scene:

- Explicit HDR reference (PQ uncompressed)
- Hidden reference (PQ uncompressed)
- Hidden reference (uncompressed HLG)
- compressed versions of PQ using HEVC at 2.5, 5, 10 and 20 Mbit/s
- compressed versions of HLG using HEVC at 2.5, 5, 10 and 20 Mbit/s

Before Test 1, an instruction sheet was given to participants (see Annex A § A.1). It was explained that for each sequence they watched, they had to assess the overall video quality for the entire duration. In addition, the visibility level of visual degradations could be used to assess the video quality.

For Test 1 on HDR content (DSCQS) viewers had to assess the following test conditions for each sequence:

- Reference (PQ uncompressed used in every A-B-pair but not marked as reference)
- Hidden reference (uncompressed HLG)
- compressed versions of PQ using HEVC at 2.5, 5, 10 and 20 Mbit/s
- compressed versions of HLG using HEVC at 2.5, 5, 10 and 20 Mbit/s

Before each test session a short verbal introduction was given to the assessors (see Annex A § A.3).

For Test 2 on SDR content (SAMVIQ only), viewers had to give their opinion on the perceived video quality of the following SDR sequences:

- Hidden reference (SDR “native” uncompressed)
- Hidden reference (HLG-SDR)
- compressed versions of SDR “native” using HEVC at 2.5, 5, 10 and 20 Mbit/s
- 4 compressed versions of HLG-SDR using HEVC at 2.5, 5, 10 and 20 Mbit/s

As in Test 1, an instruction sheet was given to participants (see Annex A § A.2).

4.6 Test methods to evaluate subjective picture quality

HDR technologies and their use in video is rather recent therefore traditional assessment methodologies need to be carefully studied and adapted to provide repeatable and consistent results. Several assessment methodologies exist for video quality assessment: in the context of subjective video quality evaluation but no methodology is preferred or even standardised for HDR.

The test labs decided to use two well-known methods: “*Double Stimulus Continuous Quality Scale*” DSCQS (Rec. ITU-R BT.500) and “*Subjective Assessment Method for Video Image Quality*” SAMVIQ method (Rec. ITU-R BT.1788) for their evaluation. Both methods use a continuous quality scale and results can be easily compared.

4.6.1 DSCQS (Test 1 only)

In DSCQS all sequences are presented in a time sequential order A-B-A-B while A is the reference and B is the test sequence or vice versa (example in Figure 4).

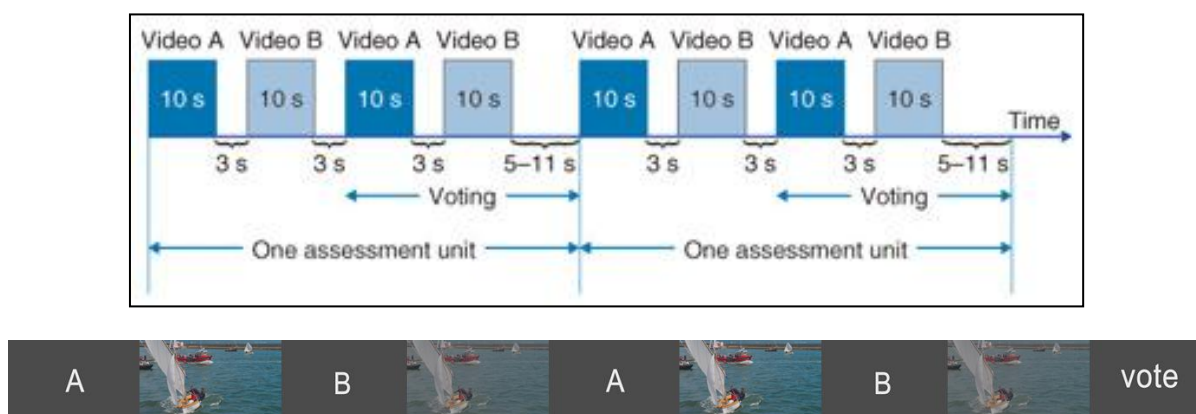


Figure 4: Example of time sequential order in DSCQS method

A test session consisted of a training part in the beginning followed by a set of test sequences. Each session is limited to a maximum of 30 minutes duration. Three sessions were performed in the environment shown below in Figure 5.



Figure 5: Viewing room for DSCQS at IRT

The PQ-based uncompressed sequence was used as the reference. Additionally, a hidden reference has been introduced - the uncompressed HLG material. All subjects marked their scorings on the continuous scale. The votes were done on paper to avoid any light sources during the viewing, as illustrated in Figure 6.

Subjective Test

Name: Gruppe Nr.: ... Session: 1 Sitz (ankreuzen):

Datum: Zeit:

	Beispiel	1 Giro	2 Zuerich	3 Football	4 Gasquet	5 Gasquet	6 PetiBato	7 4ever_short
	A B	A B	A B	A B	A B	A B	A B	A B
excellent	 	 	 	 	 	 	 	
good	 	 	 	 	 	 	 	
fair	 	 	 	 	 	 	 	
poor	 	 	 	 	 	 	 	
bad	 	 	 	 	 	 	 	

Figure 6: Sample form of the scoring sheet for DSCQS

4.6.1.1 Infrastructure for DSCQS-tests (IRT)

The play out of the content to be tested was carried out using a video server (Rohde & Schwarz Clipster) that can handle uncompressed UHD content (DPX, 10-bit, 4:2:2 sub sampling) and both transfer functions in bypass mode (transparent pass-through of ST.2084 and HLG). The server was directly connected to the Sony BVM-X300 monitor using 4x 3G SDI.

Due to the missing information in the SDI (which transfer function is currently played back) a manual switching of the transfer function in the monitor was needed without disturbing the viewer.

This has been realised using the Ethernet Remote Control of the BVM-X300 and a time accurate switching of the transfer function using an EDL⁴. The infrastructure is shown in Figure 7. The switching happened within every “grey” between A and B (see § 4.6.1) although it was not needed if both A and B were ST.2084.

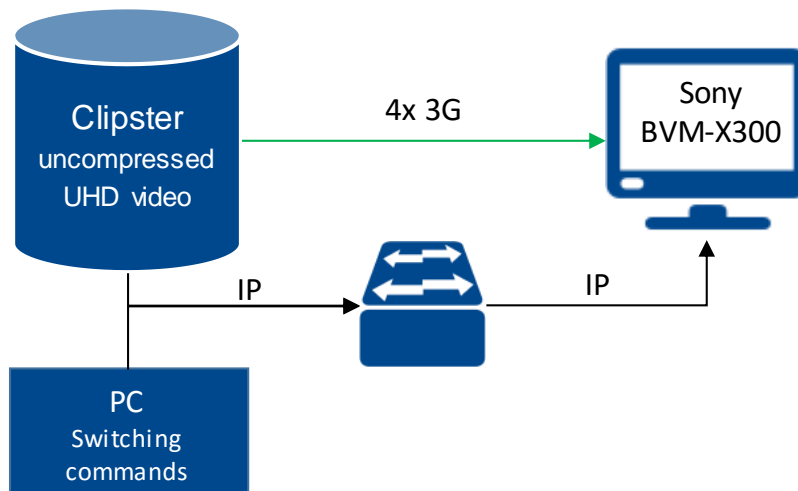


Figure 7: DSCQS Test infrastructure

4.6.2 SAMVIQ (Test 1 and Test 2)

4.6.2.1 Introduction

In the SAMVIQ methodology (Recommendation ITU-R BT.1788), each viewer has access to all sequences and the explicit reference (if it exists) at any time and in any order the viewer wants. The viewer can watch each sequence as many times as he/she wants. This methodology improves the discrimination between the different cases that have to be scored. The scoring is done using a continuous quality scale graded from 0 to 100 and annotated with 5 quality labels: Excellent, Good, Fair, Poor and Bad (Figure 8).

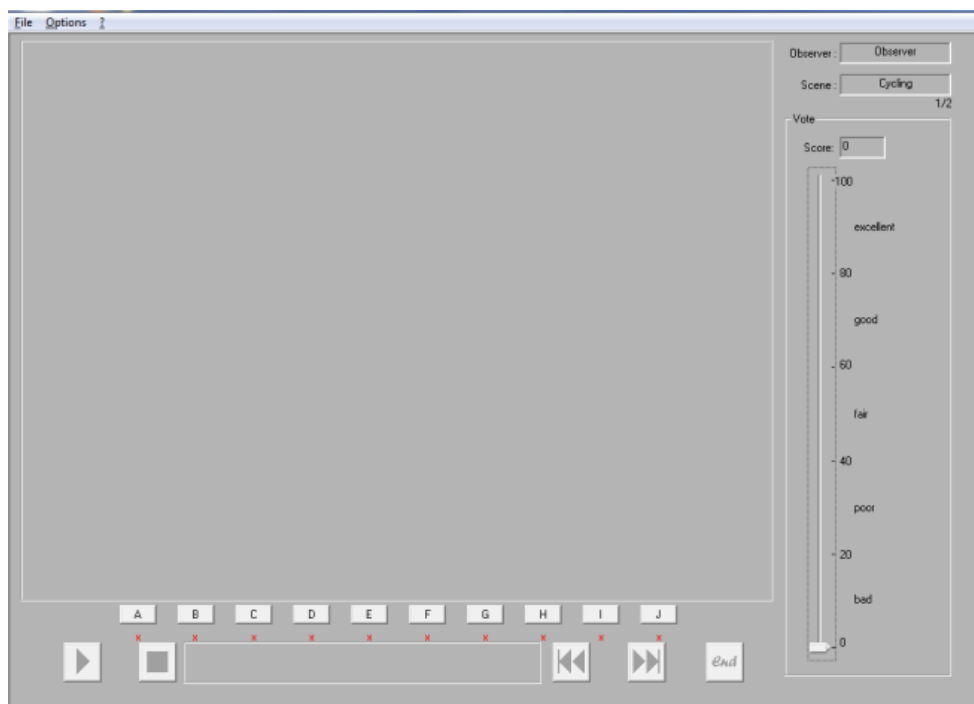


Figure 8: SAMVIQ user interface

⁴ EDL - Edit Decision List: Timecode information for IN and OUT of each clip



Figure 9 Testing room at Orange Labs

4.6.2.2 Infrastructure for SAMVIQ-tests (RAI and Orange)

The test bed architecture is composed of 4 main components (Figure 10). The Seovq software manages the subjective test using a specific user interface compliant with the SAMVIQ protocol.

All test sequences are stored in YUV420 10-bit video format on the Video Player PC. Each YUV file is associated to a configuration file describing the video format as well as the display features to change. Thus, the video player can configure the Sony display with the correct parameters (contrast level, transfer function, etc.) adjusting “User Preset” and “Input Settings” using the Ethernet interface of the display.

When the presentation of a video sequence is required from the Seovq PC, the Video Player PC plays the sequence using four 3G-SDI interfaces directly connected to the Sony BVM-X300 display.

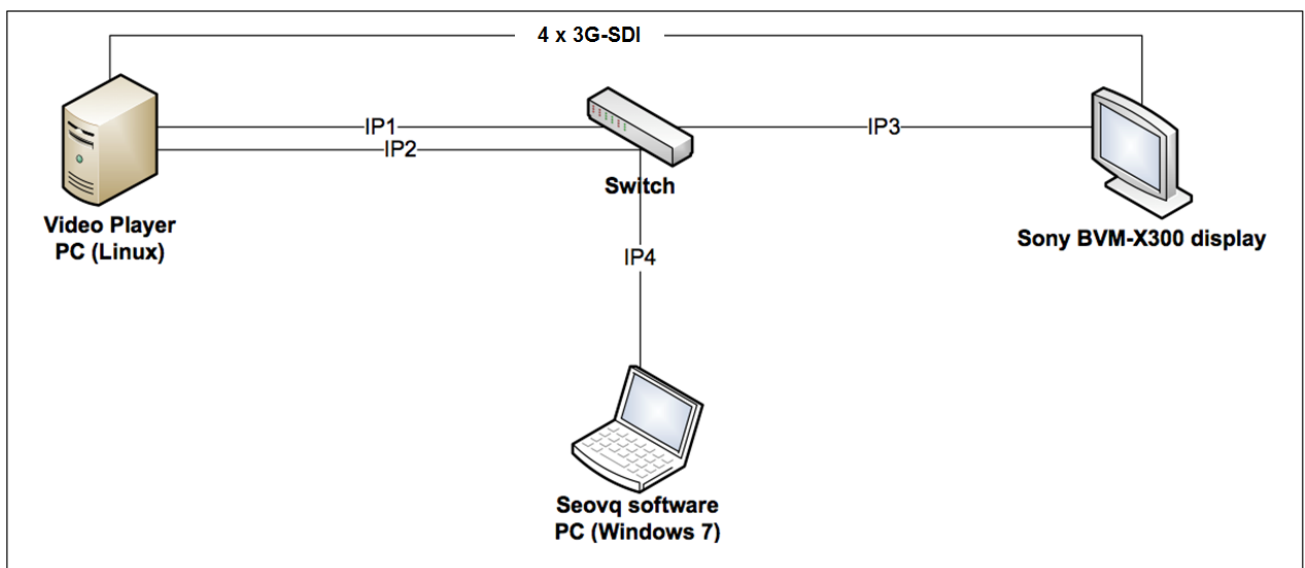


Figure 10: Test bed architecture

5. Test results

5.1 DSCQS

Following are the results of the DSCQS tests performed at IRT. About 20 viewers participated in the test and all were included in the final results (no rejections).

Table 1: DSCQS-results for Test 1

Arrangement	Cycling	Fireworks	Football	Movie	Sailing	Tennis	Overall
HLG2.5	57.9	63.6	77.85	62.45	46.6	71.4	63.3
95% conf	12.588	16.609	8.404	12.107	15.487	10.012	5.378
HLG5	72.2	73.35	96.35	83.85	61.25	90.45	79.575
95% conf	9.291	10.295	5.671	6.460	12.401	4.401	4.169
HLG10	94.2	88.45	97.25	87.8	78.1	97.25	90.508
95% conf	5.269	5.775	2.865	6.951	8.626	3.649	2.732
HLG20	95.75	92.6	98.7	95.5	79.05	97.15	93.125
95% conf	4.414	6.266	4.266	4.831	7.797	2.359	2.528
HLG Ref	100.75	100	100.5	99.05	98.35	103.15	100.3
95% conf	2.563	2.304	1.884	2.756	3.076	4.404	1.246
PQ Ref	99.6	98.8	97.9	99.6	98.45	99.55	98.983
95% conf	3.366	2.882	4.2	2.796	2.496	1.759	1.241
PQ2.5	58.1	55.95	74.35	59.85	46.95	67.1	60.383
95% conf	13.531	13.696	11.695	12.615	15.448	10.344	5.388
PQ5	69.2	72.75	89.55	77.3	61	91.65	76.908
95% conf	10.934	10.849	9.423	8.025	12.391	6.217	2.914
PQ10	84.6	88.05	95.15	96.5	76.5	95.65	89.408
95% conf	6.348	9.239	3.598	5.649	8.885	4.717	3.096
PQ20	96.8	87.85	97.45	97.1	73.05	98.05	91.717
95% conf	4.03	6.713	2.662	3.792	9.332	3.797	2.914

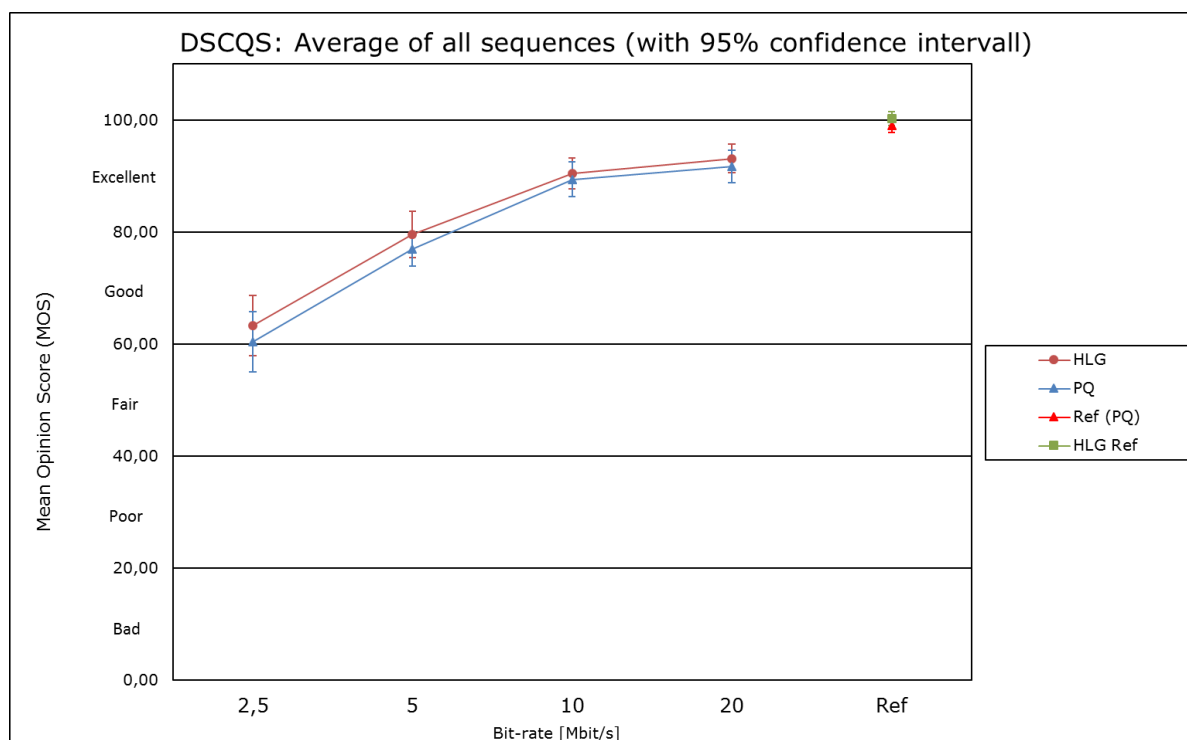


Figure 11: DSCQS: Global results for all sequences (Test 1)

The overall DSCQS results have shown that there is an equivalent subjective quality for both transfer functions (HLG and PQ) in the encoded domain. There are some differences in some sequences (e.g. cycling at 10 Mbit/s or movie at 10 Mbit/s) but still with an overlapping of the confidence intervals. All detailed results per sequence are listed in Annex B.

5.2 SAMVIQ

The tests carried out at RAI and Orange Labs provided similar results. It was decided to merge the results from the two laboratories.

34 and 35 people participated in Test 1 and Test 2 respectively. Considering the rejection criteria described in the Recommendation ITU-R BT.1788, 30 people were finally selected for results analysis of each subjective test.

Test 1 and Test 2 results statistical analyses are provided below (Table 2 and 3). For each test condition, the mean value and the confidence interval are provided.

The graphs related to global results (over all scenes) are presented in the next section and graphs related to each scene are provided in Annex B.

Table 2: Results of test 1 (HDR)

Scene Name	Type	2.5 (PQ)	5 (PQ)	10 (PQ)	20 (PQ)	REF (PQ)	2.5 (HLG)	5 (HLG)	10 (HLG)	20 (HLG)	REF (HLG)
Cycling	AV	25.3	44.1	55.7	75.4	85.3	21.3	44	62.8	77.7	81.6
	CI	5.5	6	5.9	3.3	4.2	5.2	6.2	5.3	3.8	3.6
Fireworks	AV	25.2	49.1	51.7	63.7	84.7	25.5	42.8	55.2	69.6	79
	CI	5.2	6.6	5.6	5.9	3.6	5.8	6.7	6	4.8	4.6
Football	AV	38.3	58.2	70.4	78.6	80.8	53.7	62.7	75.1	76.9	76.2
	CI	6.4	7.5	7.2	4.9	4.9	7	6.3	5.2	5.3	5.2
Movie	AV	27.2	49.9	68.9	82.5	83.8	27.4	56.8	74.9	81.5	84.2
	CI	6.2	7.1	4.8	3.8	3.5	6.2	6.2	4.9	4.5	3.8
Tennis	AV	44.7	70.7	74	75.9	81.5	48.3	76.1	82.7	80.9	82.9
	CI	6.8	5.7	6.7	6.1	3.8	7.2	4.3	3.9	4.5	4.5
Sailing	AV	23	41.5	52.8	59.6	87.7	21.4	45.6	63.7	72.7	89.1
	CI	4.7	6.6	7.1	7.9	3.1	4.6	6.6	6	5.6	2.8
Global	AV	30.6	52.2	62.3	72.6	84	32.9	54.7	69.1	76.5	82.2
	CI	2.6	3	2.9	2.5	1.6	3.1	3	2.5	2	1.8

AV = Average, CI = Confidence Interval

Table 3: Results of test 2 (SDR)

Scene Name	Type	2.5 (SDR)	5 (SDR)	10 (SDR)	20 (SDR)	REF (SDR)	2.5 (HLG)	5 (HLG)	10 (HLG)	20 (HLG)	REF (HLG)
Cycling	AV	33.2	48.8	70.4	79.6	78	34.3	52.7	70.1	77.4	77.9
	CI	6.4	6.7	4.6	3.5	4.6	6.7	6.1	3.5	5.1	3.2
Fireworks	AV	38.6	51.5	64.1	71.2	76.2	28.2	51.3	65	71.9	80.3
	CI	6.1	6.8	5.9	5.2	5.2	6	6.1	6.5	4.2	3.7
Football	AV	49.7	61.7	70.3	67	73.5	47.6	61.9	69.8	67.8	70.9
	CI	7.1	6.4	3.8	5.5	6.1	6.2	7.3	5.4	6	4.2
Movie	AV	39	58.9	65.7	66.4	79.3	35.1	66.1	79.2	81.7	68.4
	CI	7	6.5	7.3	6.6	4.2	6.7	6.4	3.7	3.4	8.2
Tennis	AV	57.2	72.1	77.1	76	77.6	54.5	68	70	75.8	78.3
	CI	6.7	4.3	4.4	3.8	3.4	6.9	6	4.7	4.9	3.8
Sailing	AV	25.7	38.4	78.1	75.9	85.4	31.4	53.7	72.1	79.8	83.3
	CI	6	6.2	5.3	4.9	3	6	6.9	5.5	3.8	4.6
Global	AV	40.6	55.2	70.9	72.7	78.3	38.5	59	71	75.7	76.5
	CI	3	2.9	2.3	2.1	1.9	2.9	2.8	2.1	2	2.1

AV = Average, CI = Confidence Interval

Global results for Test 1 (HDR) carried out at RAI and Orange Labs are presented in Figure 12. These show that HLG-HDR provides equivalent or slightly better Mean Opinion Scores than PQ10.

From a statistical point of view, a Student t-test carried out on each bitrate level and for each scene show that HLG-HDR is at least equivalent or better than PQ10 (except for *Fireworks* at 5 Mbit/s).

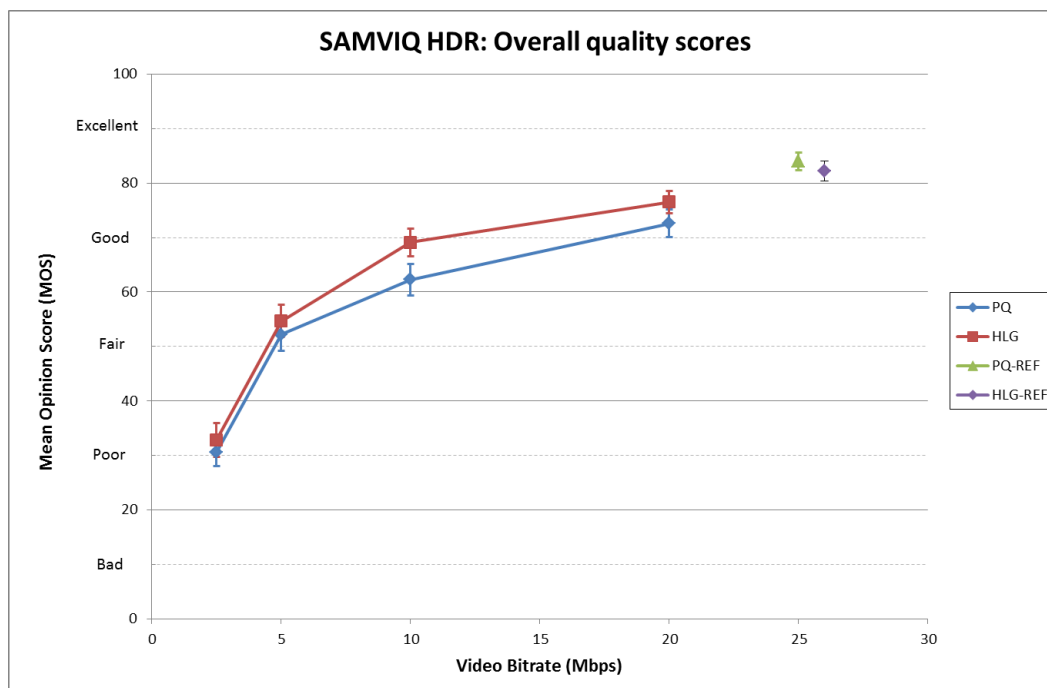


Figure 12: Global results over all scene (HDR)

Concerning the PQ to HLG-HDR conversion (performed on the uncompressed source material), it can be noticed that this process does not impact the perceived quality of original source content for Sailing, Movie, Tennis and Cycling. For Football and Fireworks contents, the statistical analysis shows better Mean Opinion Scores for PQ10.

Global results for Test 2 (SDR) carried out at RAI and Orange Labs are presented in Figure 13.

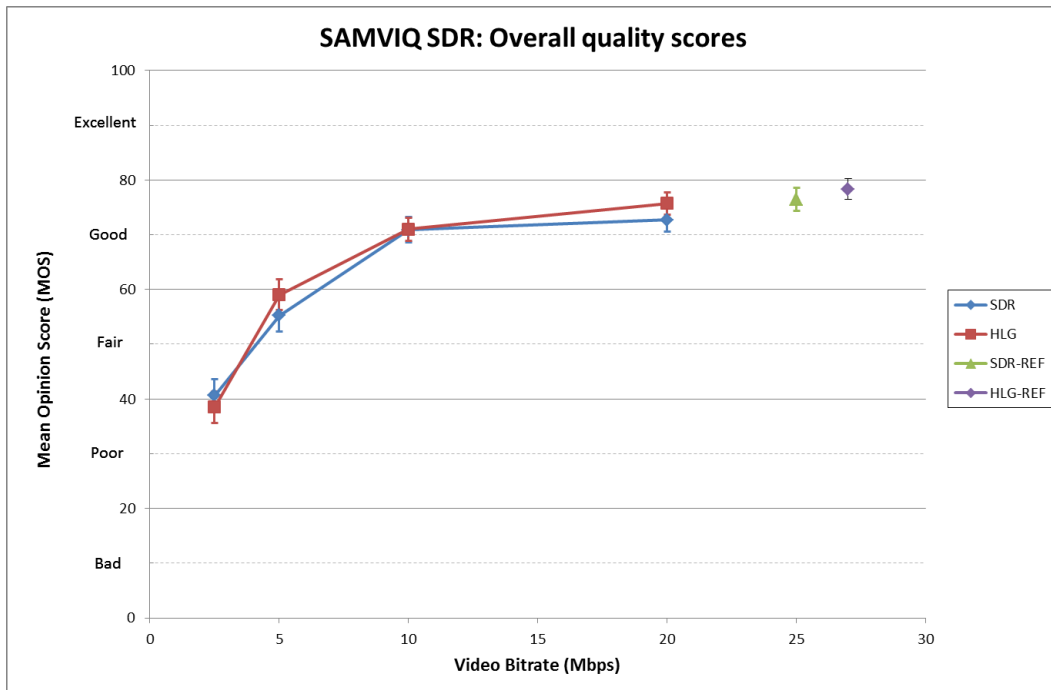


Figure 13: Global results over all scene (SDR)

Results show that HLG-SDR and “SDR manual grade” provide equivalent perceived video quality after compression. A Student t-test carried out on each bitrate level and for each scene show that HLG-SDR is at least equivalent than “SDR manual grade” (except for *Fireworks* at 2.5 Mbit/s and *Tennis* and *Sailing* at 10 Mbit/s).

Concerning uncompressed source contents, the perceived video quality is equivalent for both SDR and HLG-SDR.

6. Main Conclusions

The tests have shown that both HDR formats (HLG10 and PQ10) chosen within DVB have the capability to provide equivalent visual quality for HDR in distribution. Even in the uncompressed domain both signals are comparable. Both methods are valid to provide HDR with a certain bit-rate to the users’ home.

HLG10 and its SDR backward compatible proportion fulfil an equivalent visual quality to a SDR native signal.

The tests were performed at 1000 cd/m² because this is the current target value for TV production peak luminance. For displays with higher peak luminance capabilities the performance of the HDR technologies should be tested considering these new conditions.

Annex A: Instructions to participants at the tests

A.1 Instruction for viewers (SAMVIQ Test 1)

Instructions

Welcome to Orange Labs / RAI,

You are about to take part in an evaluation of the quality of video sequences (video only, no sound). For each sequence, you have to assess the overall video quality for the entire duration. In addition, the visibility level of visual degradations can be used to assess the video quality.

Six various clips of about 10 - 15 seconds long have been selected (football, tennis, etc.). Each of them has been treated with different processes indicated by the letters A, B, C, etc.). The reference clip ("REF" button) has not been processed.

You may view each sequence in any order and repeat it as many time as you want (at least one time entire duration) using the "Play" button. After the visualization of each sequence, you can report your opinion moving the slider on the quality scale (numbered from 0 to 100) according to quality labels "Bad", "Poor", "Fair", "Good" and "Excellent".

The scoring can be modified or refine at any time. You have to score the sequences of one clip before to assess the next clip pressing the "Fast Forward" button.

At the end of the last sequence of the last clip, the "END" button becomes active. Press it to complete the test session.

Thank you for your participation.

A.2 Instruction for viewers (SAMVIQ Test 2)

Instructions

Welcome to Orange Labs / RAI,

You are about to take part in an evaluation of the quality of video sequences (video only, no sound). For each sequence, you have to assess the overall video quality for the entire duration. In addition, the visibility level of visual degradations can be used to assess the video quality.

Six various clips of about 10 - 15 seconds long have been selected (football, tennis, etc.). Each of them has been treated with different processes indicated by the letters A, B, C, etc.).

You may view each sequence in any order and repeat it as many time as you want (at least one time entire duration) using the “Play” button. After the visualization of each sequence, you can report your opinion moving the slider on the quality scale (numbered from 0 to 100) according to quality labels “Bad”, “Poor”, “Fair”, “Good” and “Excellent”.

The scoring can be modified or refine at any time. You have to score the sequences of one clip before to assess the next clip pressing the “Fast Forward” button.

At the end of the last sequence of the last clip, the “END” button becomes active. Press it to complete the test session.

Thank for your participation.

A.3 Instruction for viewers (DSCQS Test 1)

Welcome:

Welcome to our subjective tests here at IRT where we want to evaluate the subjective quality of UHD video signals including Higher Dynamic Range. The tests take place at IRT and in parallel at Orange Labs and RAI. Our goal is to compare Higher Dynamic Range candidates and additionally the bit-rates needed. The tests are according to ITU-R BT.500 methods. We will start with a short training sequence.

Explanation of test paper:

On the paper you will find two vertical line pairs called A and B. A corresponds to sequence A; B corresponds to sequence B. You will see A-B and then again A-B. In between both sequences there is a gray pattern called "A" and "B" and at the end of each A-B-A-B is a pattern called "vote". You can start your "vote" after the first A-B but latest at the "vote"-pattern.

The first pair on your voting sheet is an example of how to make your decision

On the left you see the quality descriptors. They correspond to certain perception levels of quality from excellent down to bad.

The line pairs are continuous scales and you are asked to mark with a clear line according to quality you judge for each sequence in comparison.

Training sequences:

We now do a short training session - please do not mark the results paper for this training session.

Please compare both A and B; look for differences such as blocking, loss of sharpness, differences in colour or brightness/contrast (examples shown in training sequences).

Either A or B is the reference - this changes randomly.

There might be the case that A and B look the same - that's intended.

Please vote only for the A-B pair you are currently looking at.

Any questions?

We start now with the evaluation, please do not talk. If you are ready with the first page, please go to the second one.

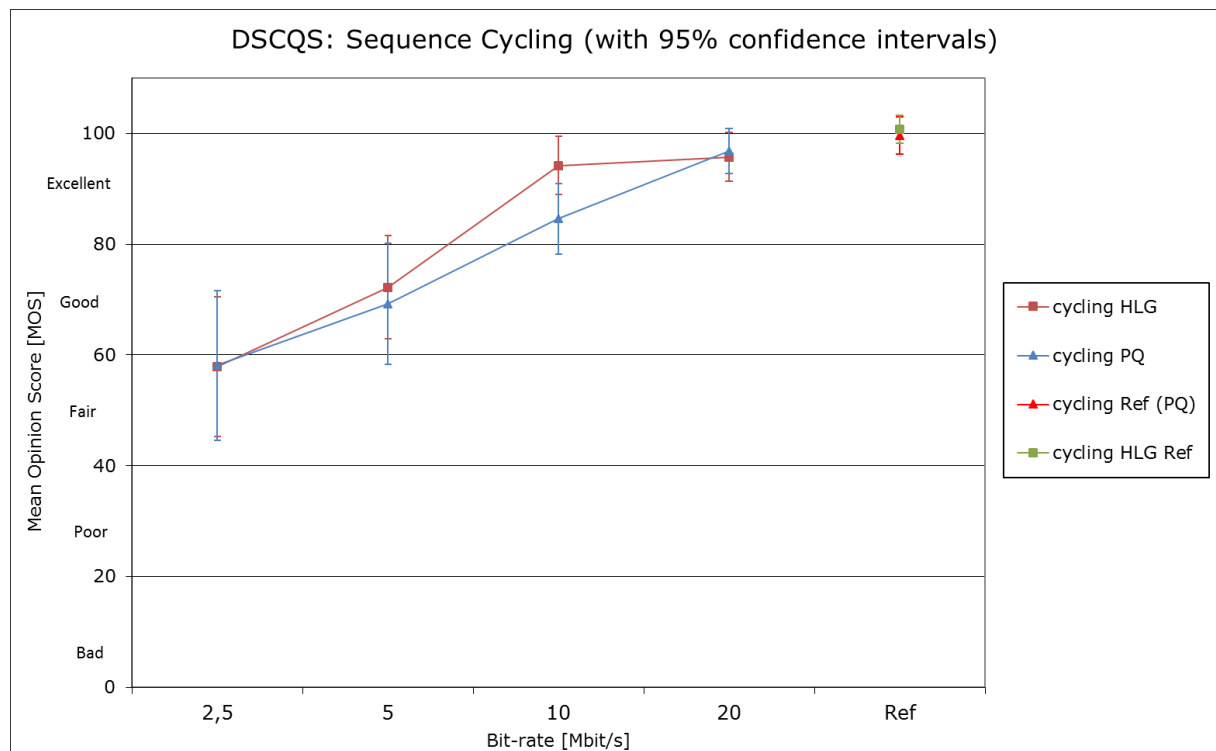
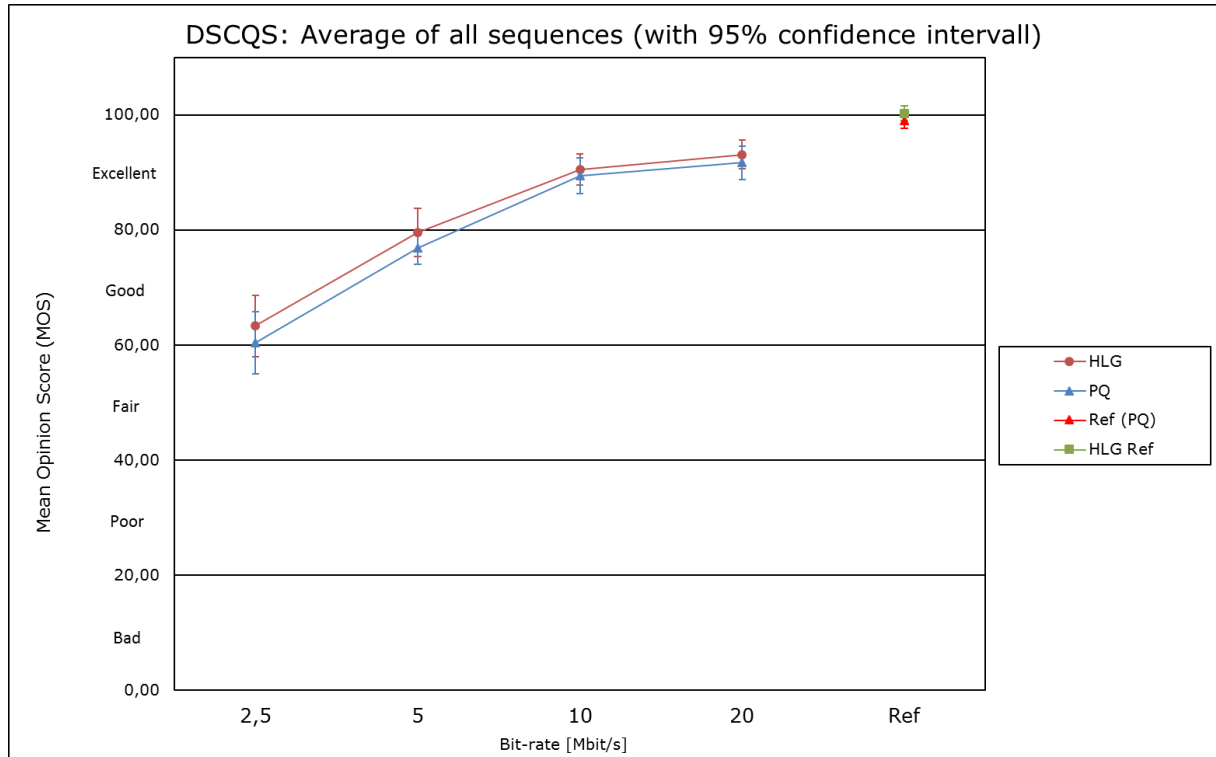
Thank you for participating in the test.

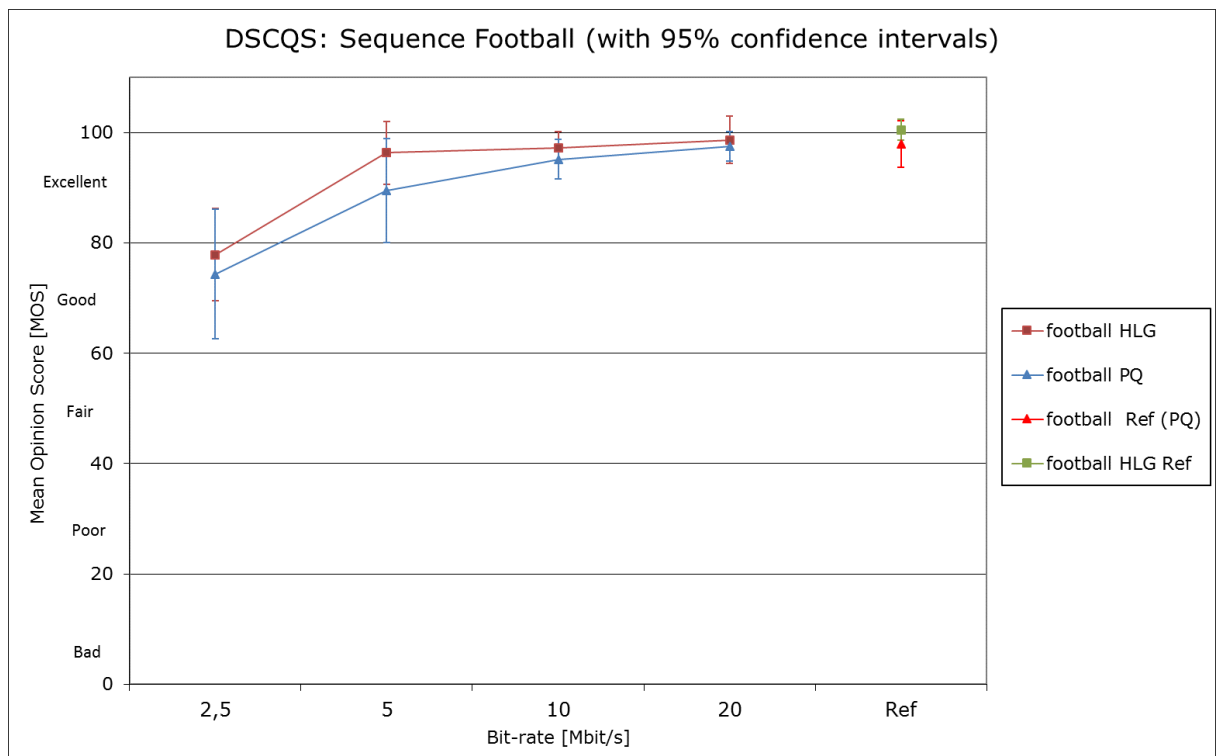
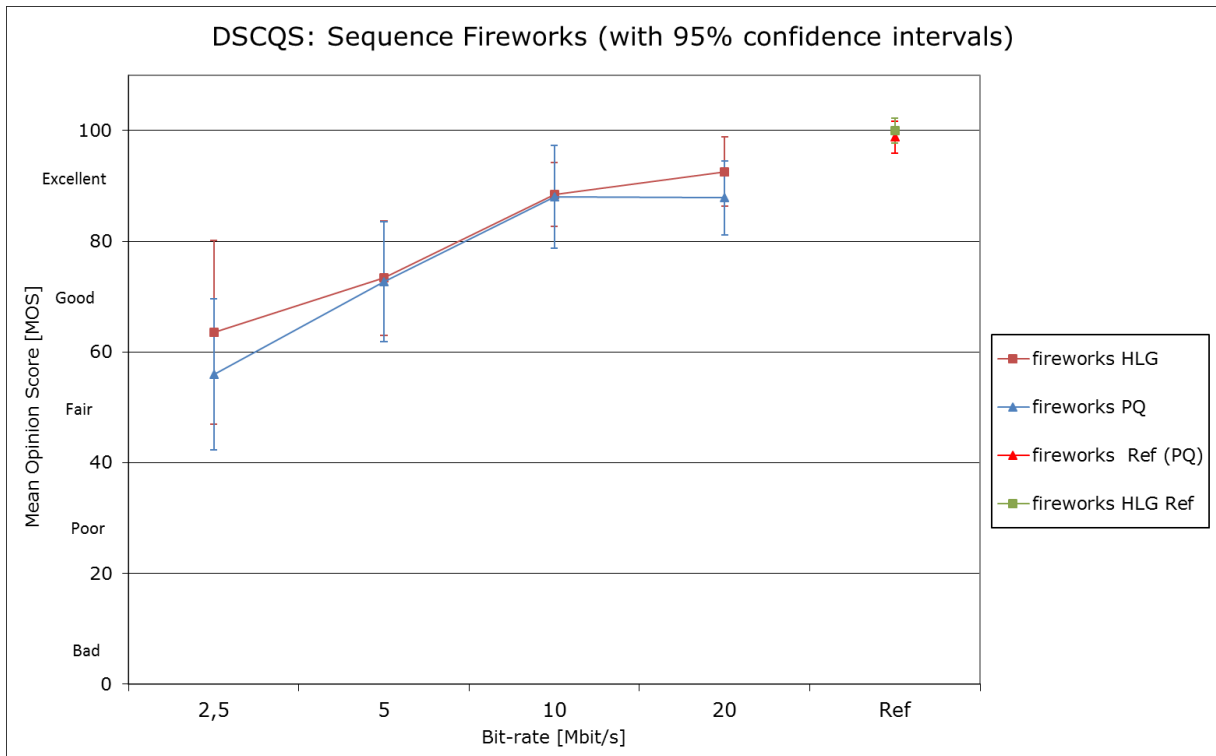
Annex B: Detailed Test Results

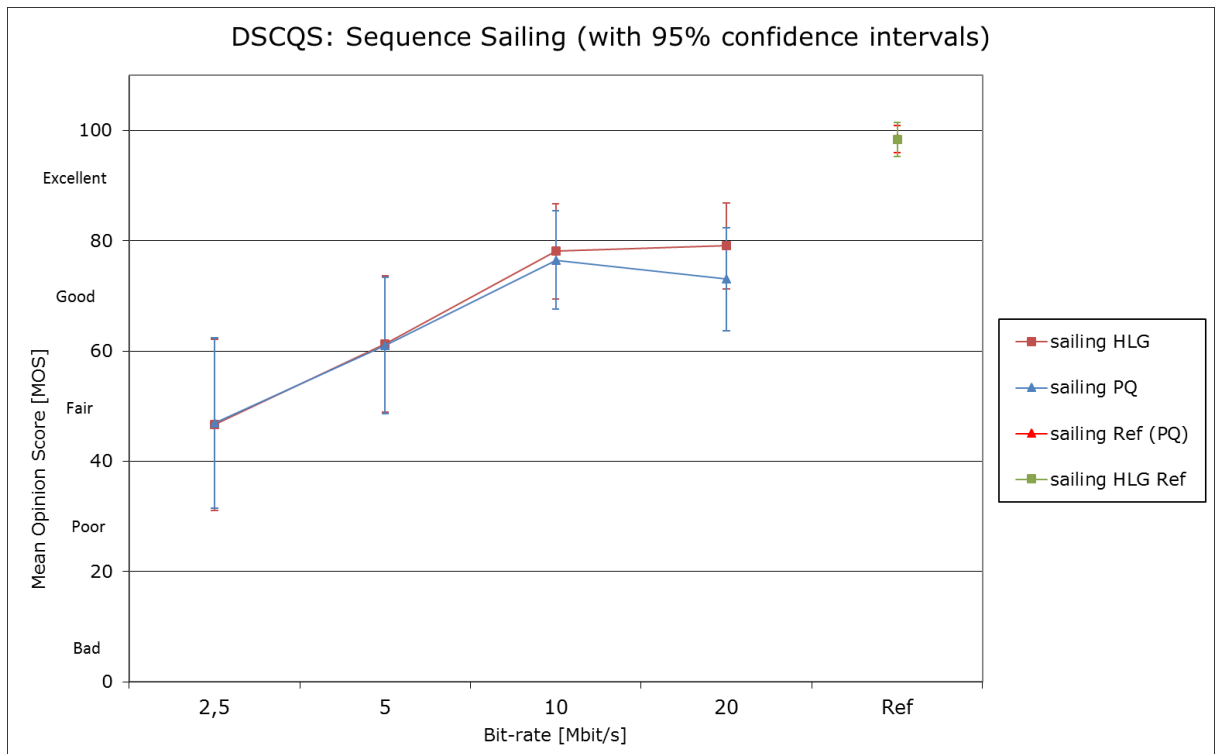
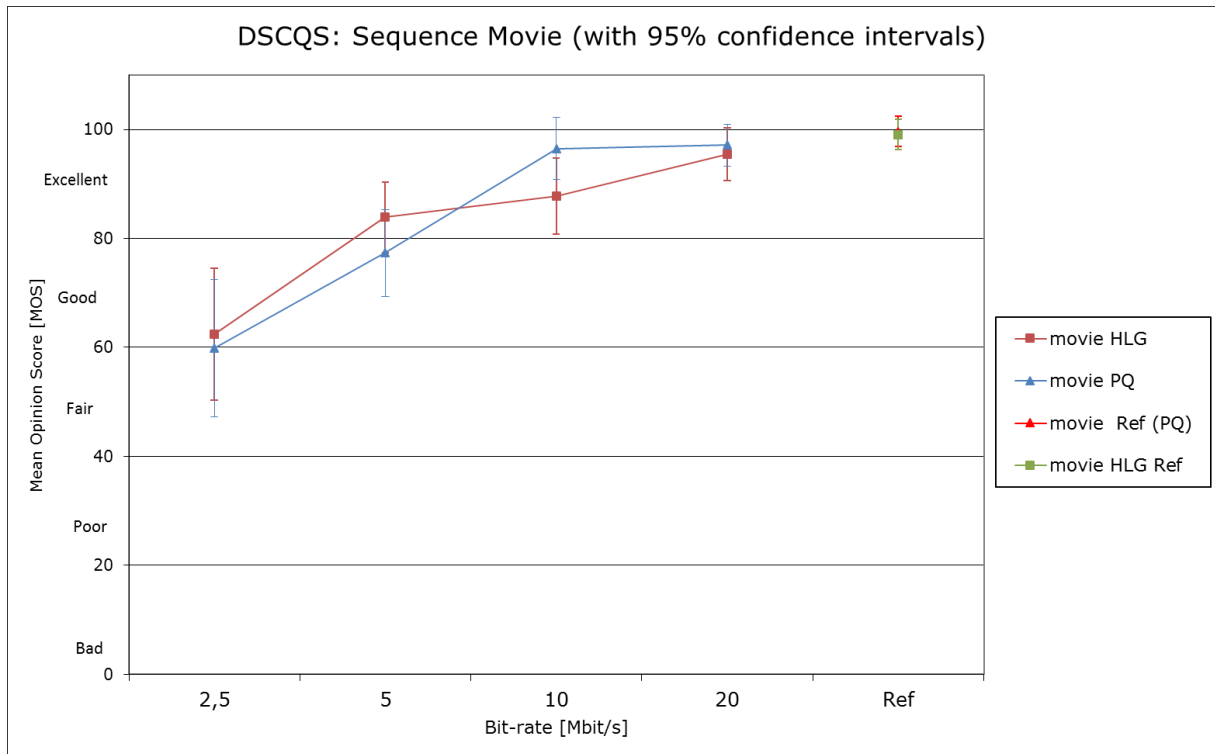
B.1 Test 1 (HDR)

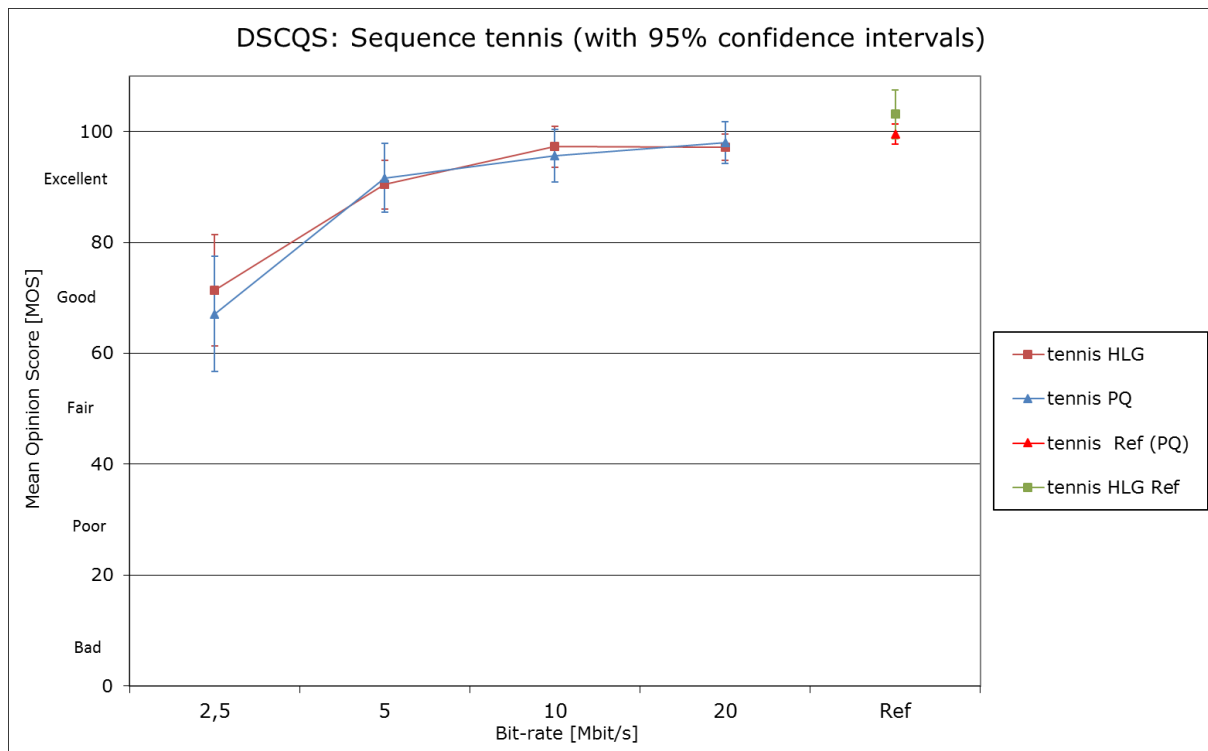
B.1.1 IRT detailed DSCQS test results for HDR

(alphabetical order of sequences).

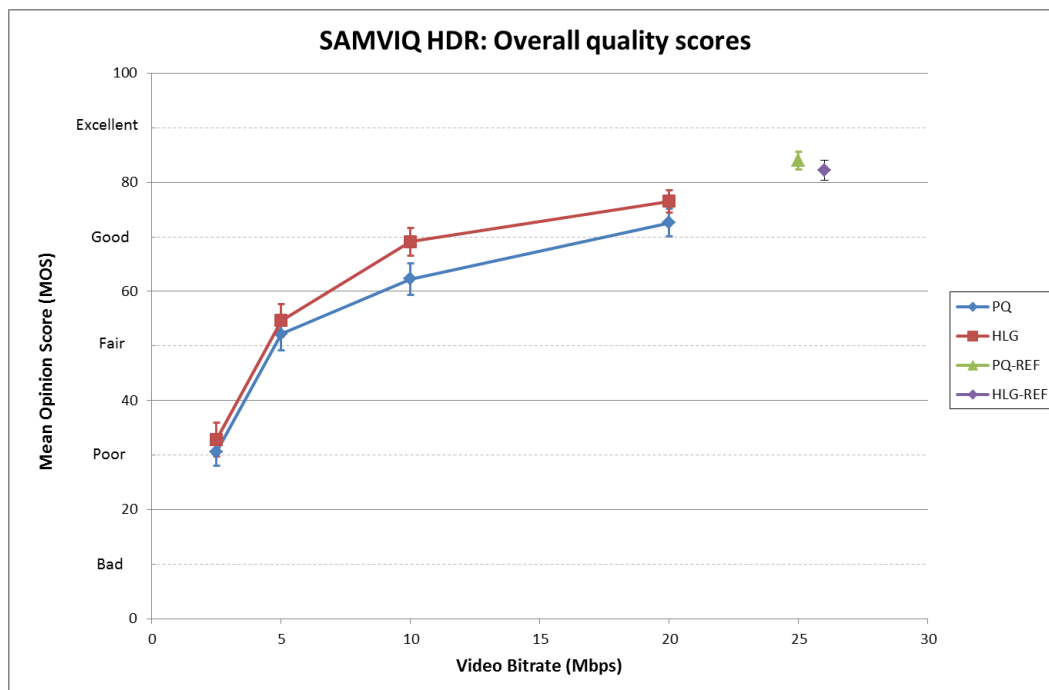


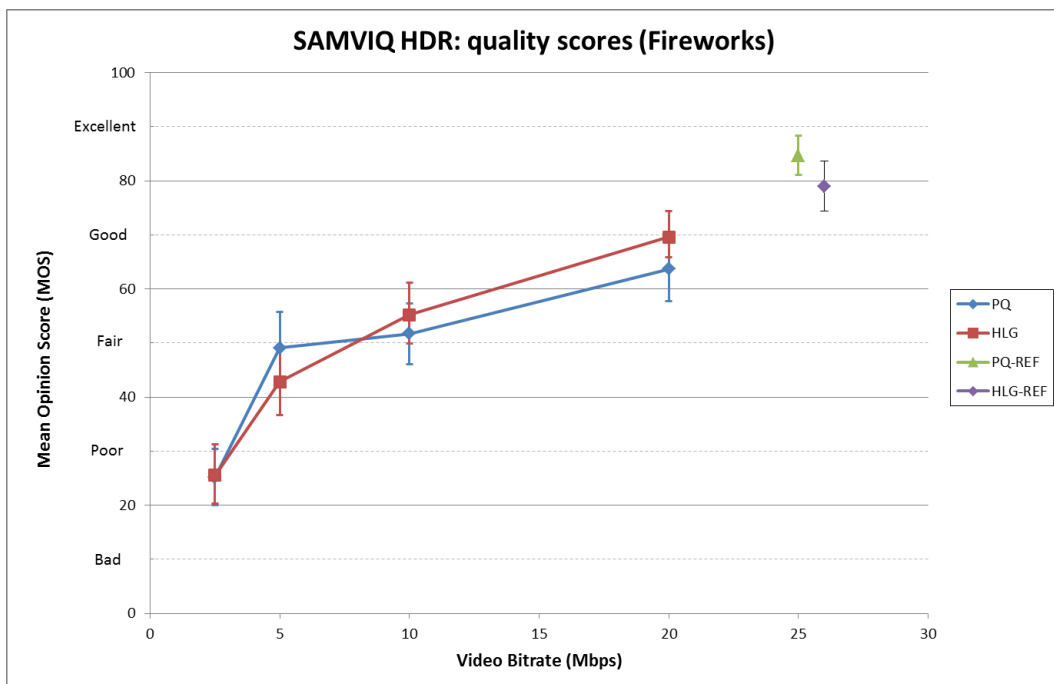
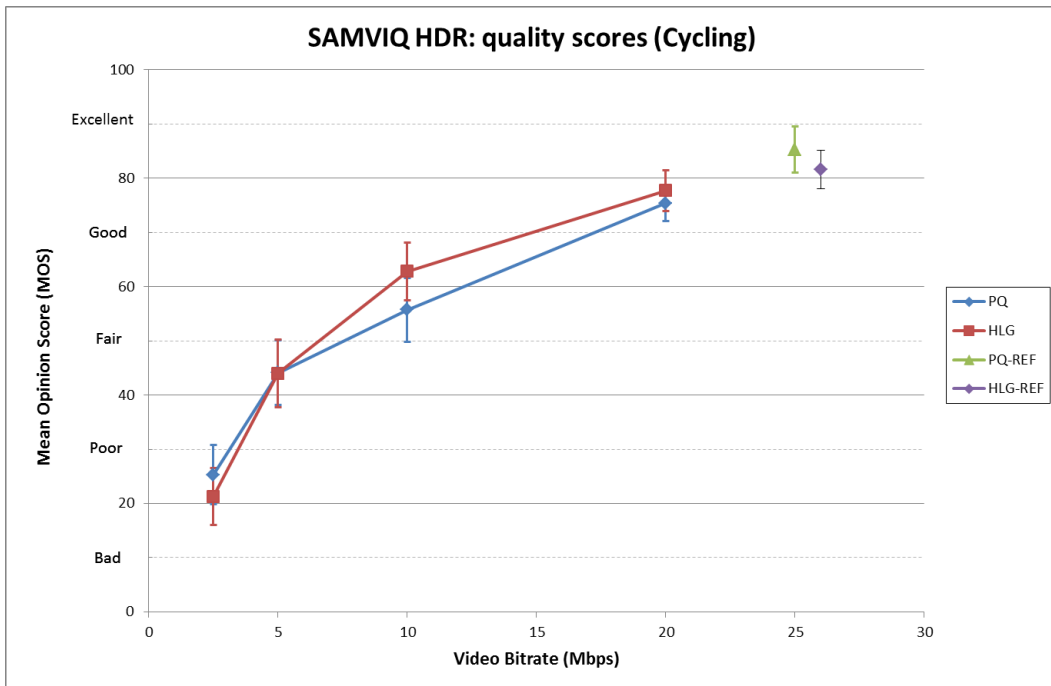


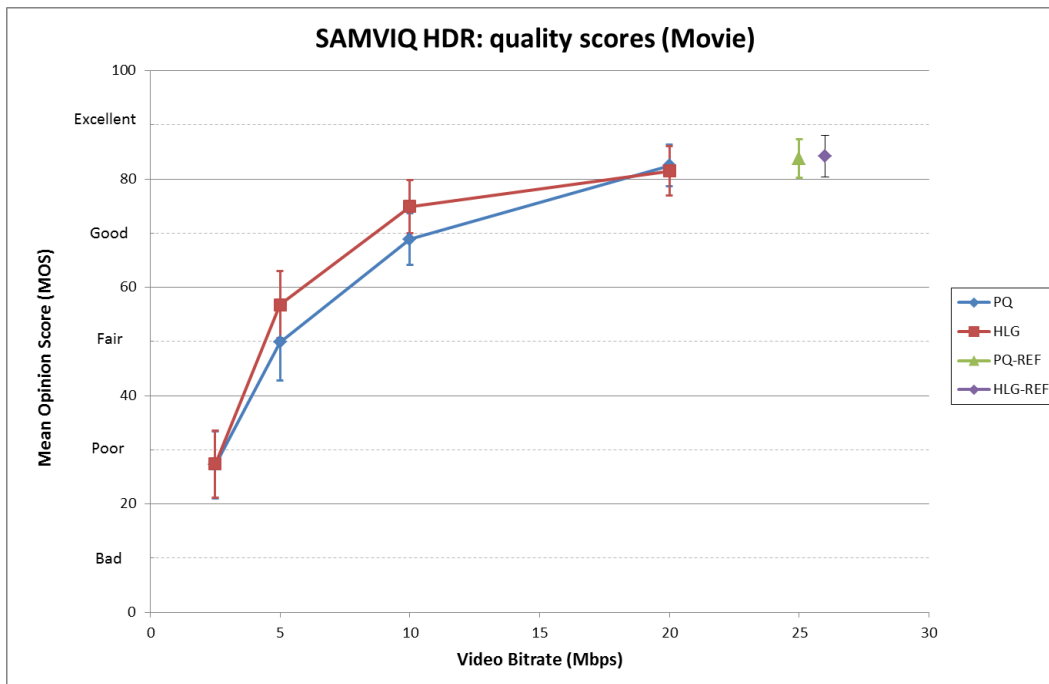
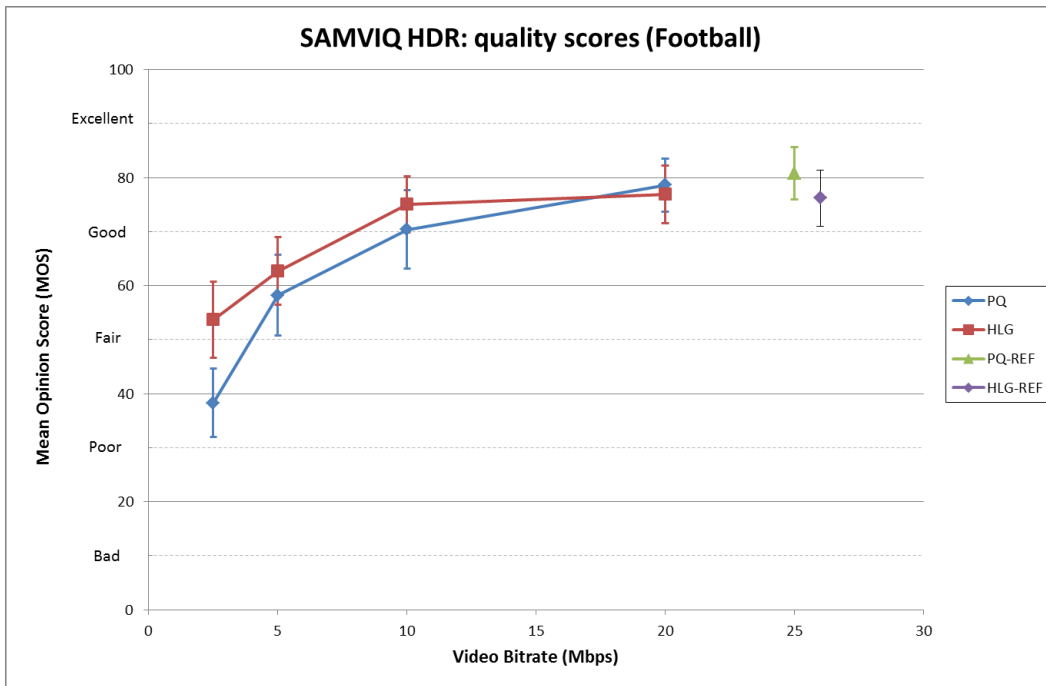


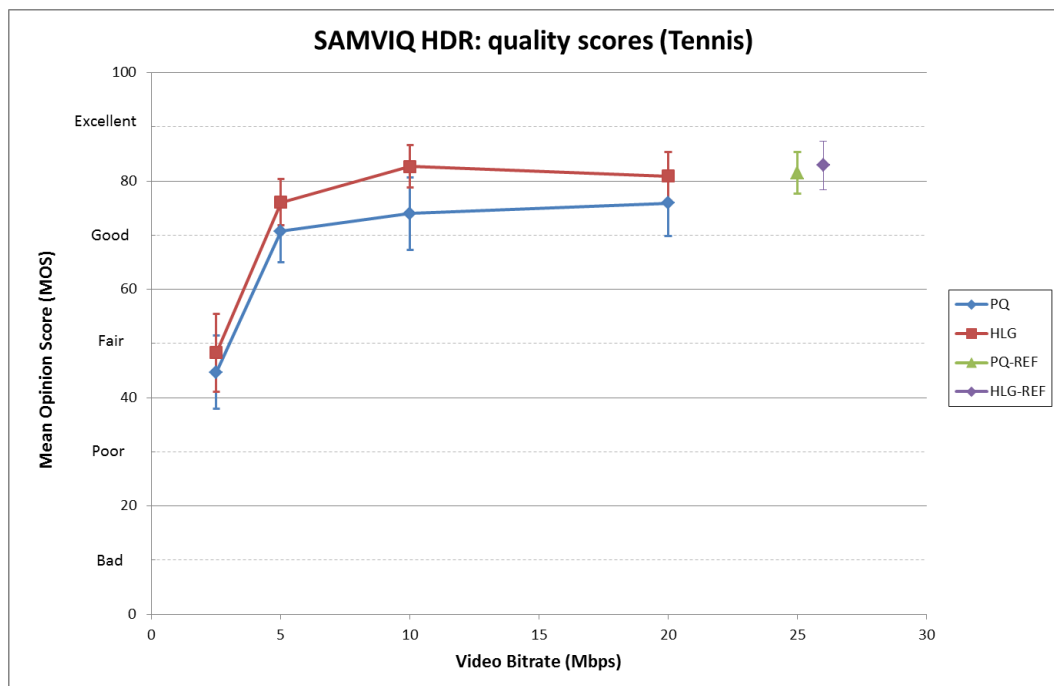
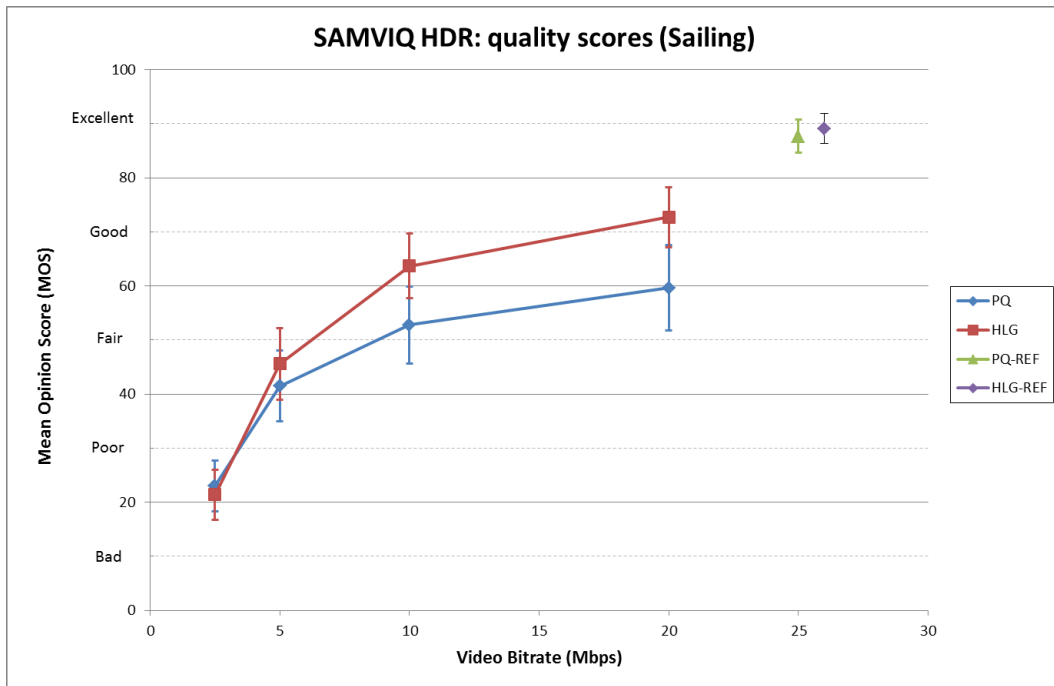


B.1.2 Orange/RAI detailed SAMVIQ test results for HDR
(alphabetical order of sequences)









B.2 Test 2 (SDR)

B.2.1 Orange/RAI detailed test results SAMVIQ for HDR

(alphabetical order of sequences)

