EBU

OPERATING EUROVISION AND EURORADIO

TR 050

SUBJECTIVE EVALUATIONS OF 100 Hz HIGH FRAME RATE (HFR)

Source: Video Systems

Geneva
August 2019

Abstract

There are various ways to improve audio-visual quality beyond current HDTV services, namely by:

1. Using more pixels (UHD);
2. Using more luminance levels (HDR);
3. Using a wider range of colours (WCG);
4. Using more frames per second (HFR);
5. Improving the audio quality (NGA).
6. And for interlaced services: moving to progressive scanning;

Public service media can choose which of the above to implement; there is no need to add all at once. The relative merits of improving the abovementioned quality aspects is studied in subjective evaluations.

This report covers two such evaluations that were performed in Q1 of 2019 at the EBU in Geneva. The goal of the evaluations was to see by how much HFR can improve image quality for various combinations of transmission format (1080p, 2160p), framerate (50p, 100p) and camera shutter (180 degrees, 360 degrees).

The results that were obtained indicate that using a state-of-the art CE display, up to 1.5 points in quality can be gained by providing 100 Hz instead of 50 Hz pictures, but that this gain can be reduced by more than half when display-side motion-compensated frame interpolation (MCFI) is considered.

Contents

Subjective Evaluations of 100 Hz High Frame Rate (HFR)

EBU Committee	First Issued	Revised	Re-issued
VS	August 2019		

Keywords: UHD, HD, HDR, HFR, shutter, blending, DSCQS.

1. Introduction

High Frame Rate (HFR) is considered as one of the key features to improve the quality beyond current HDTV. Other main features that can improve the user experience are:

- Changing from interlaced to progressive scanning.
- Using a higher definition/resolution (UHD);
- Using more luminance levels (HDR);
- Using a wider range of colours (WCG);
- Improving audio quality (NGA);

Public service media can choose which of the above to implement; there is no need to add all at once. To understand how much quality can be gained, the EBU and its Members frequently perform subjective evaluations.

This report covers two subjective evaluations that were performed in January and March 2019 at the EBU in Geneva. The goal of the evaluations was to see how much HFR can improve image quality and what the impact of several format choices is on the overall quality perceived by the viewer:

- 100 fps[1] vs 50 fps;
- For 50 fps: 180 degrees or 360 degrees shutter;
- 2160p or 1080p as the 'distribution' format.

This document describes how the tests were conducted, the results obtained, and the conclusions drawn. The tests were organized and executed by the EBU and the IRT, with assistance from the BBC and the RAI.

2. Test set-up

2.1 Test material

Fast moving content is expected to benefit most from HFR, so recordings of several athletics disciplines were used for the tests. The material was shot in 2018 during the inaugural edition of the European Athletics Championships in Berlin, where the EBU had set up a multi-camera UHD-HDR-WCG-HDR-NGA trial production.[2] All content was originated in 2160p/100, 360°shutter, BT.2100 (HLG), DPX Sampling YCbCr 4:2:2, 10-bit, narrow range. All test sequences reflect practical scenarios and shooting conditions.[3]

[1] Frames per second; the picture refresh rate.

[2] See https://tech.ebu.ch/publications/tech-i-038.

[3] Note that (portions of) the sequences with high-speed movement effectively may not use the full spatial resolution of 2160p due to the motion blur introduced during shooting.

A set of six test sequences from the EAC2018 dataset[4] was selected. Audio was not used for these tests. Figure 1 gives an impression of the type of material used.

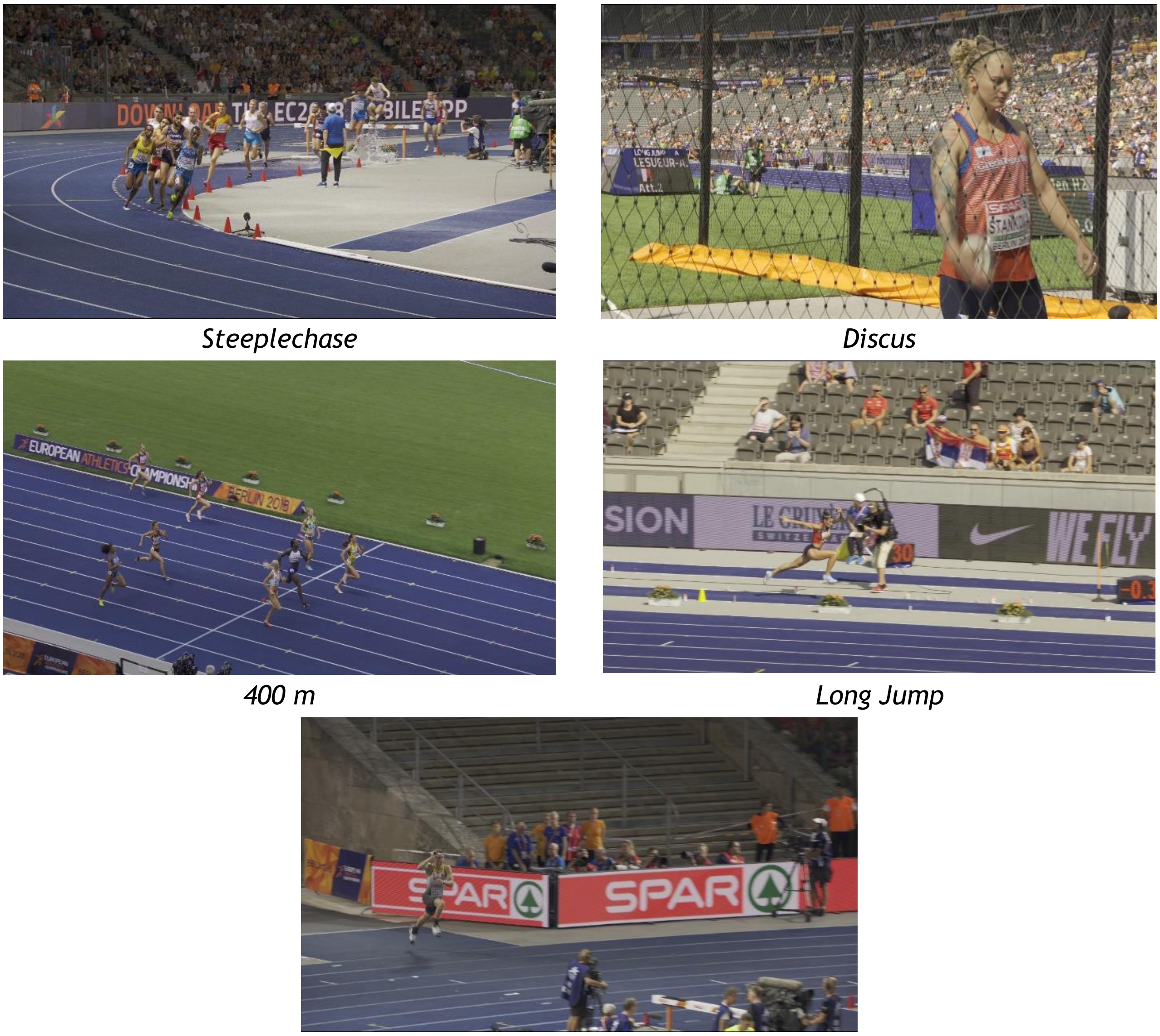

Figure 1: Keyframes from the five main test sequences used for the subjective evaluations.

Five of the sequences were used for the actual evaluation. One sequence (not shown above) was used to train the subjects at the start of each session.

2.2 Formats

For each of the sequences, six different versions of the original 2160p/100 (360° = 100% shutter) sequence were made:

A. 2160p/100 (100%) - copy of the original
B. 2160p/50 (100%) - every two subsequent original frames blended into a single new one
C. 2160p/50 (50%) - every other frame of the original sequence
D. 1080p/100 (100%) - as 1 above, but downscaled to 1080p resolution
E. 1080p/50 (100%) - as 2 above, but downscaled to 1080p resolution
F. 1080p/50 (50%) - as 3 above, but downscaled to 1080p resolution

[4] See https://tech.ebu.ch/eac2018

The sequences in format A were used as the reference.

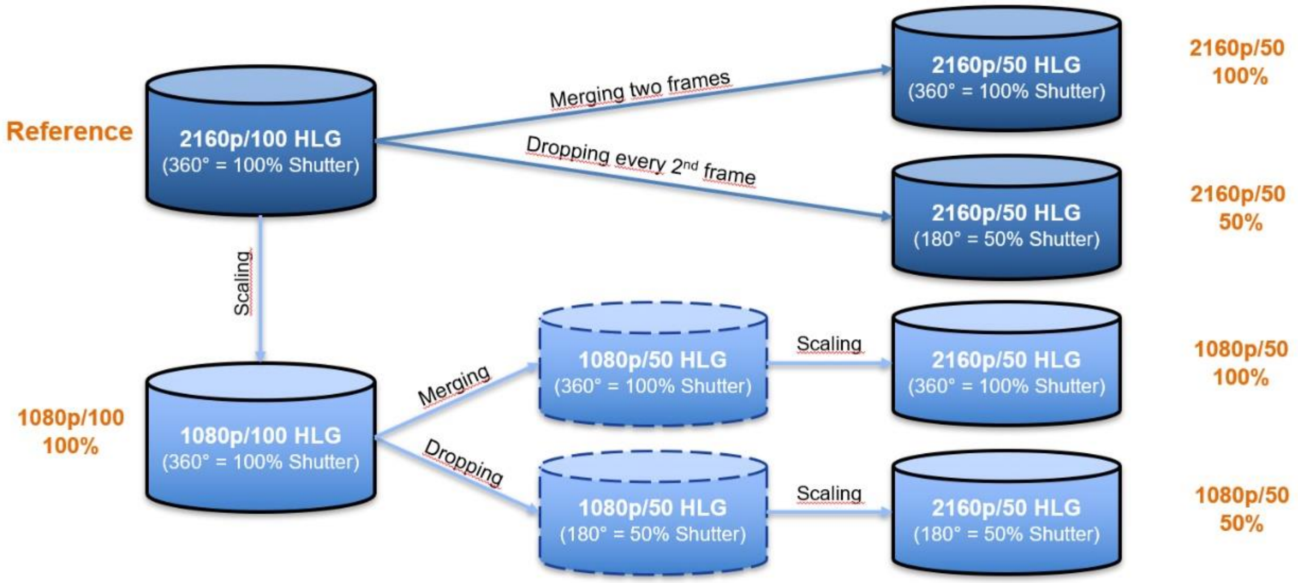

Figure 2: Overview of how the final sequences in six different formats were derived.

The spatial downscaling (2160->1080) was performed using a Rohde & Schwarz Clipster video server. The HDR settings used are shown in Figure 3.

Figure 3: HDR settings used for Clipster operations

The temporal down-sampling for the 'blended' format C (100 fps → 50 fps with a 100% shutter) was performed using a software tool kindly provided by the BBC; this tool allows the loading of video from a packed or planar format, to linearize it, accumulate and average the frames, apply gamma, and output the frames in a planar or packed format.

2.3 Play-out

For play-out during the testing sessions, all clips (see § 4) were processed as follows:

- First, all 1080p content was upscaled (using a Clipster) back to 2160p. This was to ensure that the tests were not dependent on the TV's upscaling capabilities, also allowing future subjective evaluations with these sequences using different displays.
- Secondly, all clips were (off-line) encoded with a state-of-the-art Ateme encoder, using the HEVC codec at 60 Mbit/s. This relatively high bitrate was chosen to avoid visible coding artefacts dominating the tests.
- Finally, the clips were concatenated into a single transport stream.

The resulting transport stream was modulated using a Dektec portable modulator (FEC 9/10, 27.5 MBaud for DVB-S2) and fed to a state-of-the-art 65" LG OLED TV set equipped with special firmware that made the set capable of decoding HFR and automatically detecting frame-rate changes in the transport stream.

2.3 Display

The same display and the same material were used for the January and the March evaluations. The display was a 65" LG OLED UHD-1 HDR HFR capable TV.

The difference between these evaluations was that in January the display had its motion-compensated frame interpolation (MCFI) function turned off ('True Motion' off), while in March the motion interpolation was turned on ('Smooth' pre-set), all other display settings were kept the same (all 'enhancers' off; no resolution enhancement, no de-noising).

2.4 End-to-end process

The overall end-to-end process for the most complex case (merging frames to obtain p50 material with a simulated 360° shutter) is sketched in Figure 4.

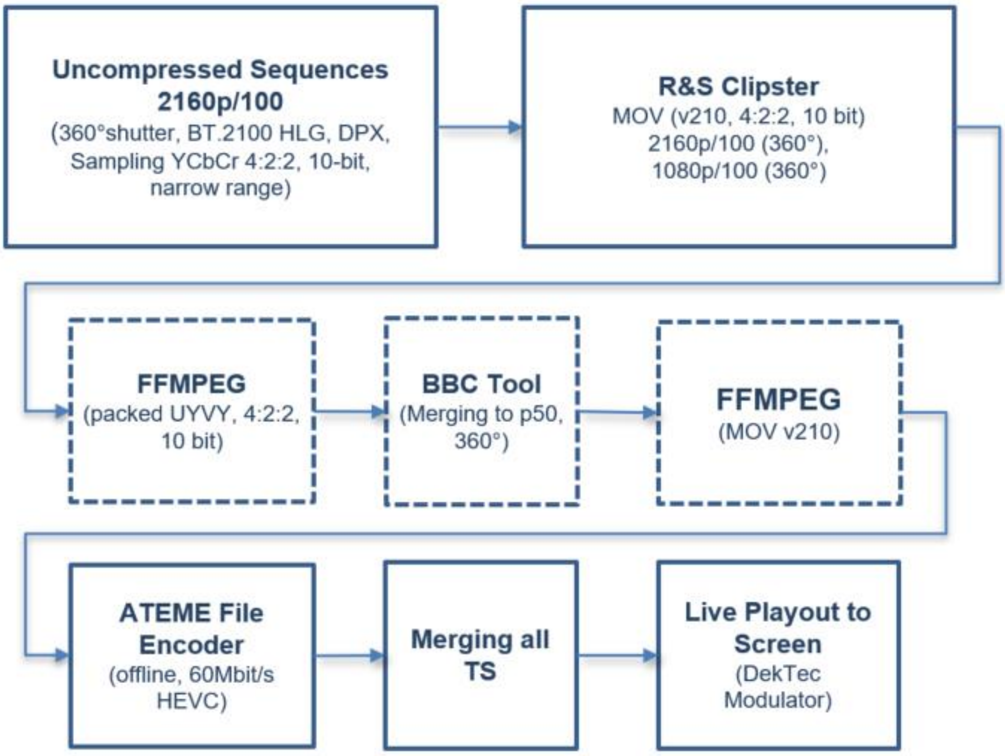

Figure 4: Overview of the main sequence preparation steps.
Note that the dashed steps were only necessary for the two blended formats (B and E).

3. Viewing conditions and viewer selection

3.1 Viewing conditions

The viewing environment respected the recommendations for critical viewing of HDR programme material defined in Rec. ITU-R BT.2100 [1].

The main characteristics of the viewing environment were:

- Viewing distance: 1.5H (= 1.5 times picture height)
- Background illuminant: D65, 5 cd/m²
- Display: LG 65" display, 100 Hz capable
- Subjects: 3 viewers per session

The viewing environments used for the January and March 2019 evaluations were comparable.

3.2 Viewer selection

The subjects comprised volunteers invited from the participants at two EBU seminars[5], together with EBU staff. Although they can be assumed to have an above average interest in, and notion of, image quality, they should not all be regarded as expert viewers.

A short verbal introduction was given to the assessors (see Annex A).

4. Test Methodology

"Double Stimulus Continuous Quality Scale" DSCQS, as defined in Rec. ITU-R BT.500[6] [2] was used to perform the evaluation.

4.1 DSCQS overview

In the DSCQS method a series of 'assessment units' is presented sequentially. In each unit one sequence-format combination is compared against the reference format[7].

The two formats are presented in a time sequential order A-B-A-B, where A is the reference and B is the format under test or vice versa. A/B can change between assessment units. At the end of each assessment unit the subject is asked to vote (score) the quality for the A and the B format individually.

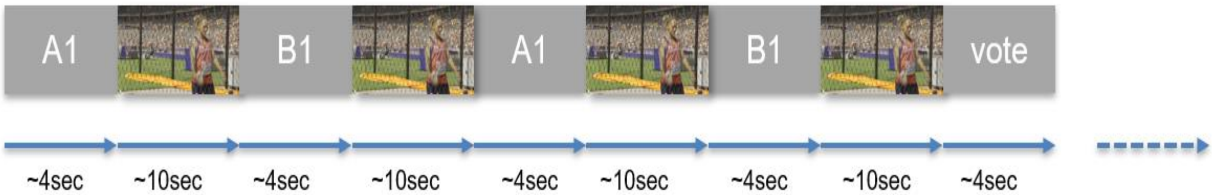

Figure 5: Illustration of the time sequential approach of the DSCQS method.
Note the number is used to indicate which assessment unit is being scored.

[5] The EBU Production Technology Seminar (29 - 31 January 2019): https://tech.ebu.ch/pts2019 , and, EBU Broadthinking (26 - 27 March 2019): https://tech.ebu.ch/broadthinking2019

[6] Although ITU-R BT.500 specifies a display peak luminance of 200 cd/m², until a specific HDR viewing environment is defined, it can be assumed valid for displays up to 1000 cd/m²

[7] Note that the sequence-format combination can be equal to the reference.

4.2 Execution

In the HFR evaluations, each test session started with a training part of two assessment units, followed by 35 assessment units. Each session had a duration of roughly 40 minutes and included up to 3 subjects.

All subjects marked their score on the 5-point continuous scale shown in Figure 6.

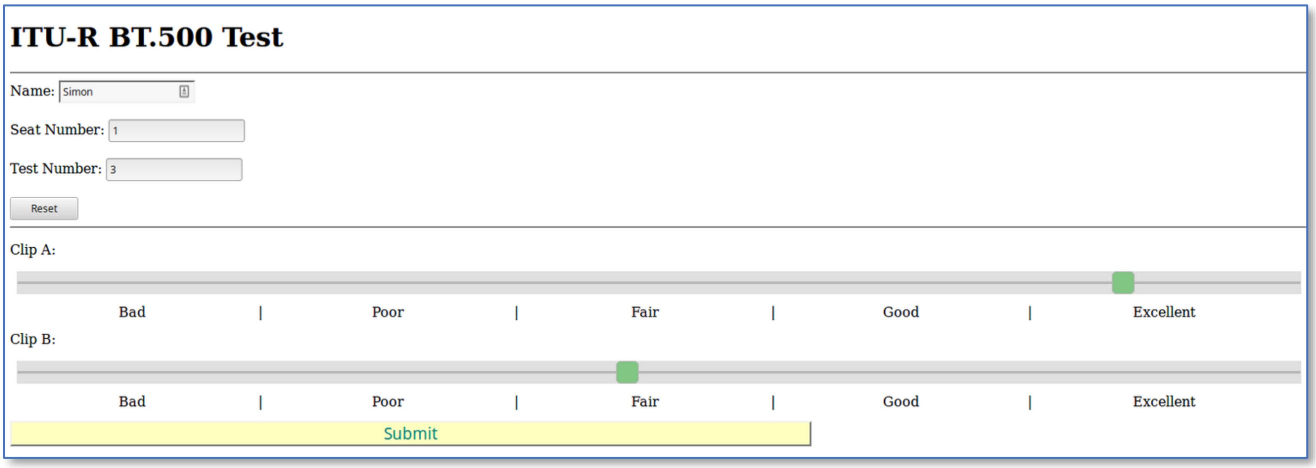

Figure 6: Tablet GUI showing the 5-point scale
The software for the scoring was developed by & and provided by BBC R&D.

Voting was done using a tablet (one per subject), which transmitted the voting values to a central database for further processing. This made performing the DSCQS analysis very efficient and avoided the need to measure scores on paper.

5. DSCQS Results

30 subjects participated in the January 2019 tests, of which 3 were discarded in the statistical analysis. 20 subjects participated in the March 2019 tests, of which 2 were discarded. The discarded observers were removed due to their large scoring difference from the mean score.

5.1 Results – no frame interpolation

The results for the evaluation performed with frame interpolation turned **off** are shown in Figures 7 and 8.

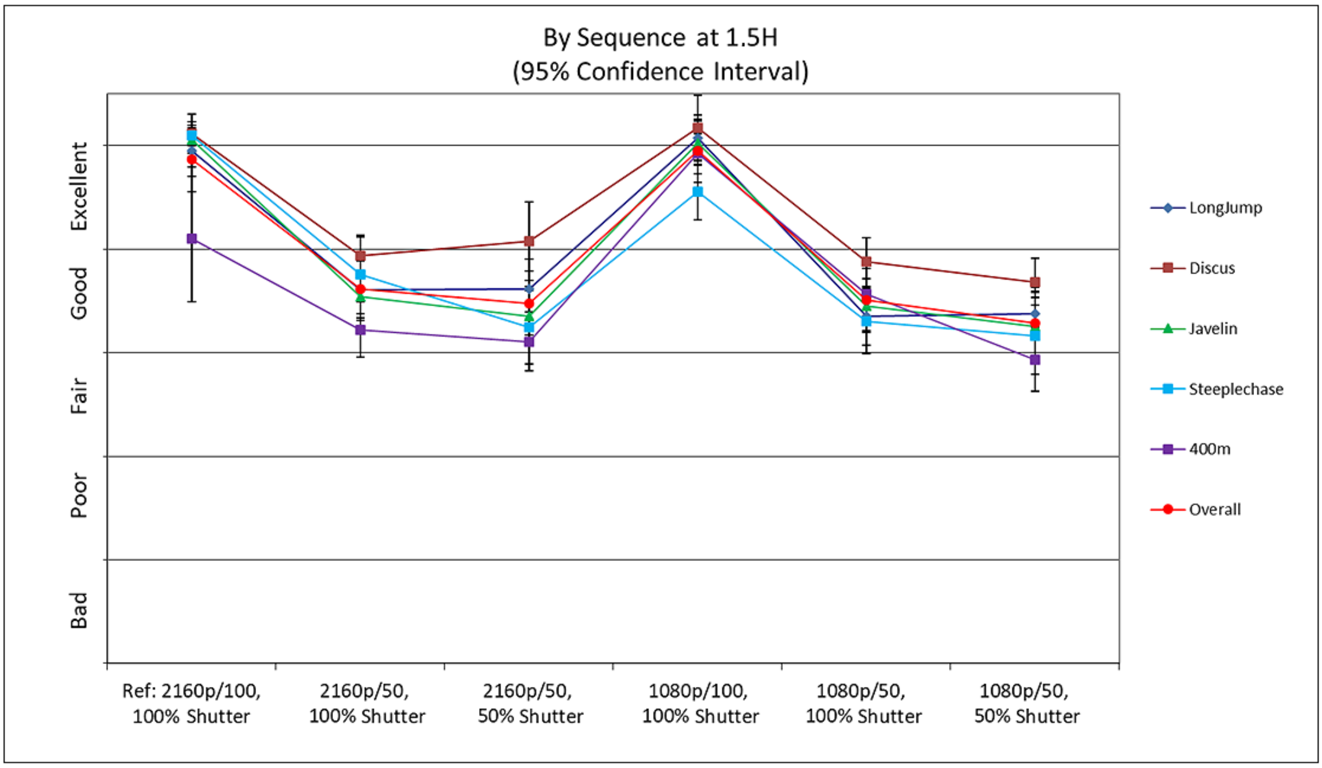

Figure 7: Results of the subjective evaluation without frame interpolation in the CE display, shown for each of the 5 sequences, including the 95% confidence intervals.

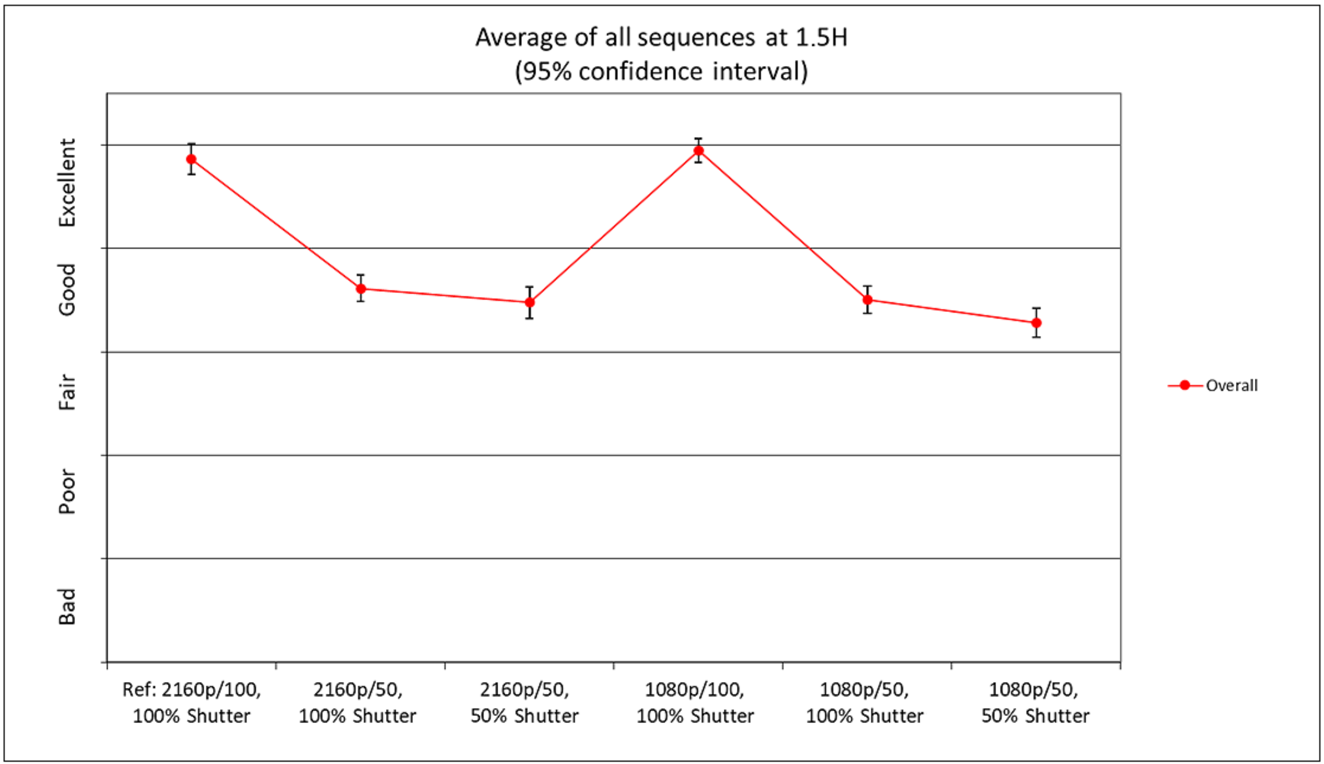

Figure 8: Results of the subjective evaluation without frame interpolation in the CE display, averaged over the 5 sequences including the 95% confidence intervals.

5.2 Results – with frame interpolation

The results for the evaluation performed with frame interpolation turned **on** are shown in Figures 9 and 10.

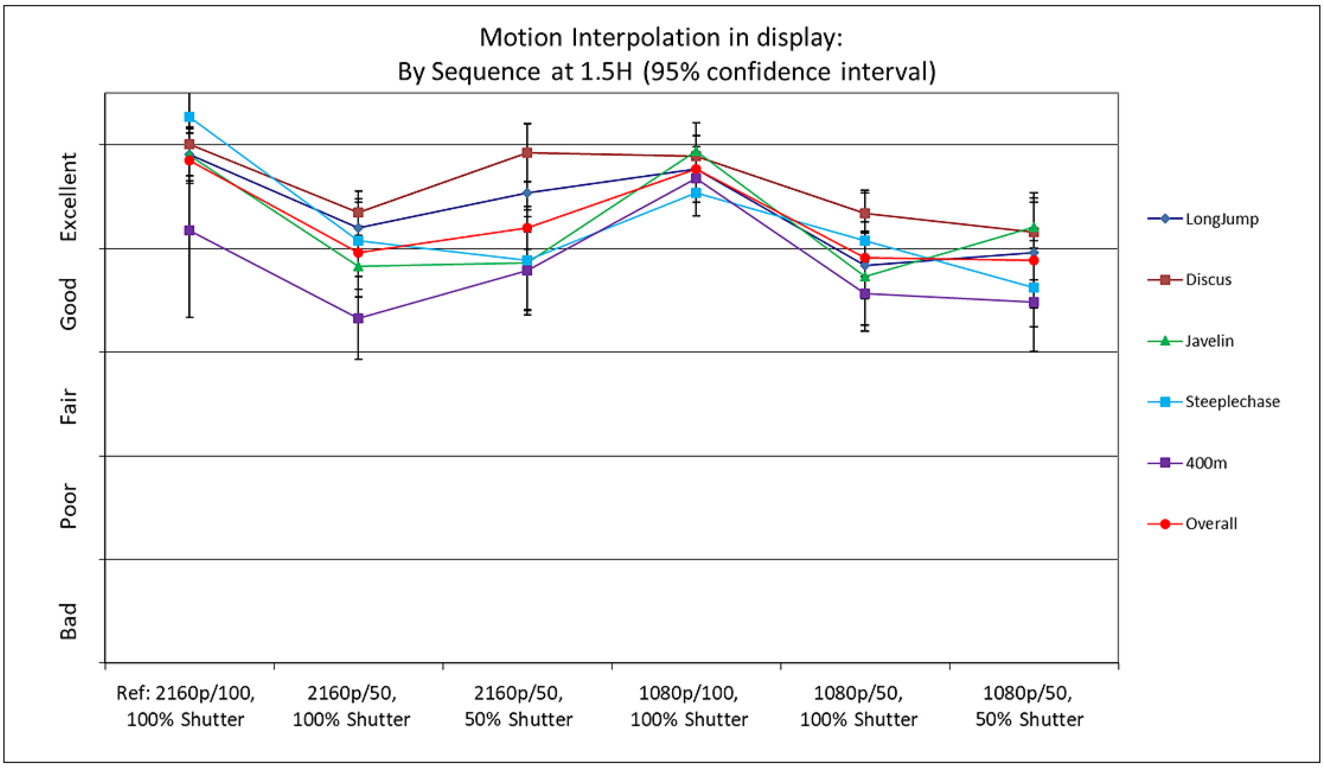

Figure 9: Results of the subjective evaluation with frame interpolation in the CE display, shown for each of the 5 sequences, including the 95% confidence intervals.

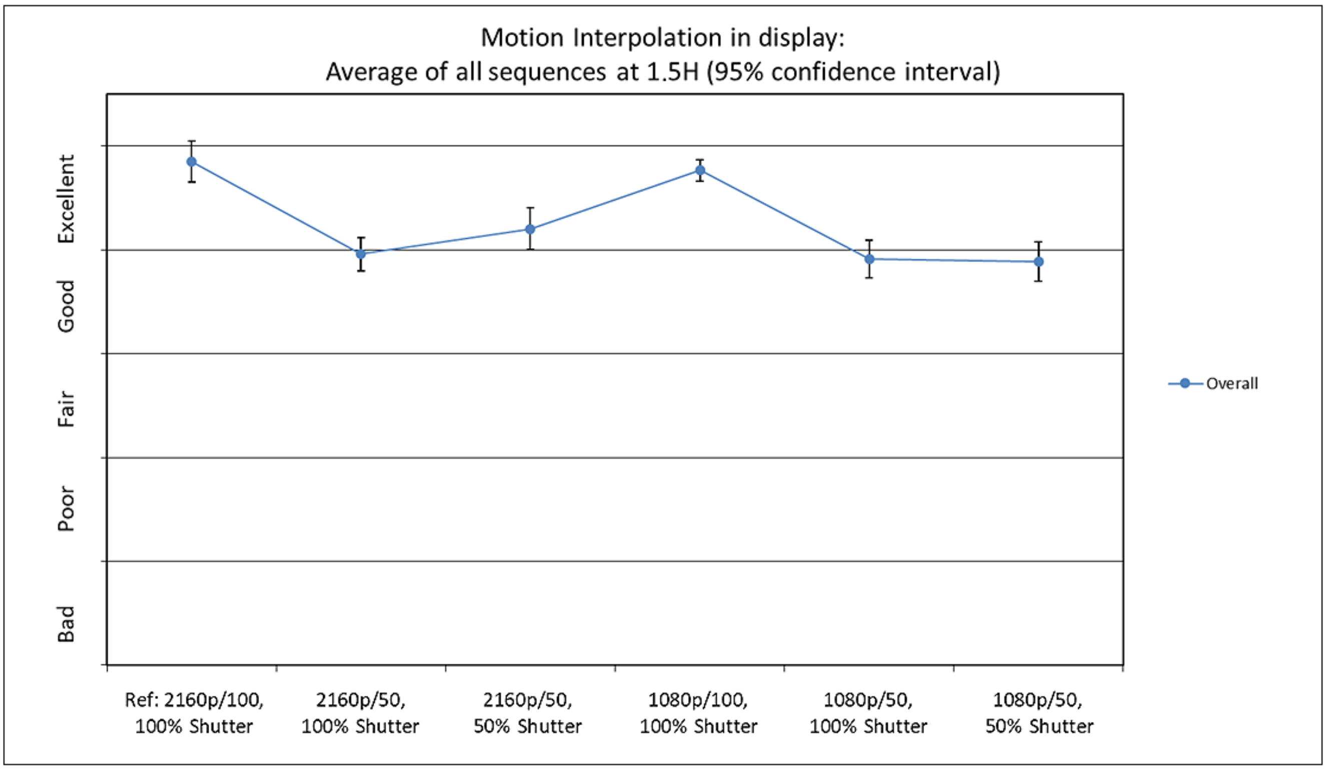

Figure 10: Results of the subjective evaluation with frame interpolation in the CE display, averaged over the 5 sequences including the 95% confidence intervals.

5.3 Results - combined

Figure 11 shows the combination of averaged results of both evaluations overlaid in a single diagram.

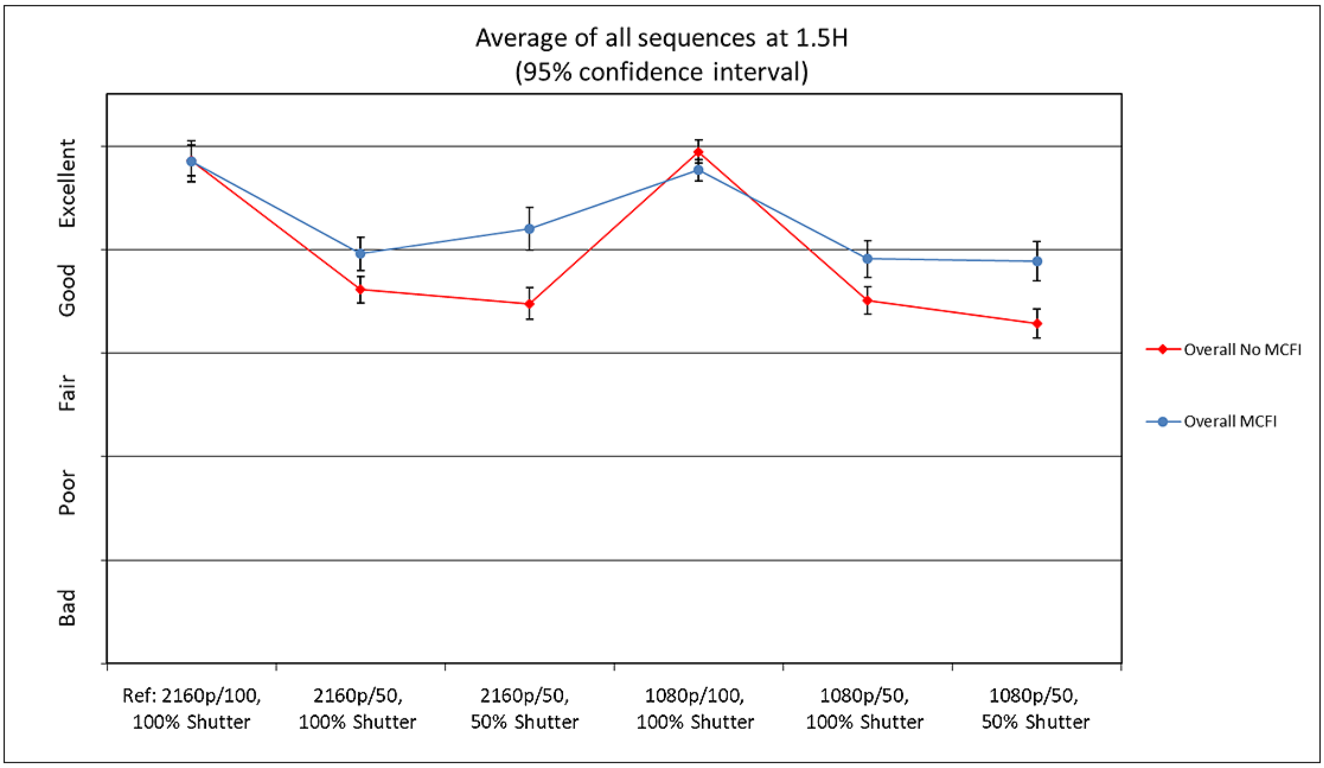

Figure 11_ Both overall results (with and without motion interpolation) and the 95% confidence intervals.

Detailed numerical results per sequence for both evaluations are provided in Annex B.

5.4 Other observations

During the testing it was observed that multiple subjects reported motion sickness, which may indicate the 1.5H distance is too close for comfort (but needed for evaluation);

The blended sequences were just perceptibly less bright in appearance than the other sequences.

6. Conclusions

HFR impact

The results indicate that without MCFI on the display side, up to about 1.5 points in quality can be gained by providing 100p instead of 50p pictures[8].

As many broadcasters are currently providing content in the 1080i/25 format, the real-world gain of introducing HFR services can thus be expected to be larger than shown in these tests – which used 50p as the 'lowest' quality.

[8] The biggest gain is 1.5 points for the 400 m sequence, when comparing the 1080p/50 50% and 1080p/100 100% formats (shown in Figure 7). Similarly, the overall average for these formats shows 1.4 points gain (illustrated in Figure 8). In both cases the worst-case combination of the 95% confidence intervals was used for the calculation (so being conservative by minimizing the gain).

Influence of resolution

The result is very similar for the 1080p and 2160p 'services' tested.

It should be noted that for both cases the material was originated in 2160p and shown on a 2160p device at the UHD viewing distance. This is in line with what sampling theory predicts and similar subjective evaluations have shown already.

Influence of MCFI

With MCFI turned on, the HFR gain may be reduced by over 50%[9].

Of course, the MCFI implementation used in these tests is only a single implementation by a single display manufacturer, so it does not provide an absolute measure, but rather a real-world indication of what state-of-the-art high-end CE displays can do. It should be kept in mind that the actual population of CE displays does not consist only of high-end models.

Material dependency

The quality gain is strongly dependent on the type of material (Figures 7 and 9).

It was expected that this would be the case, for the simple reason that the amount and the type of motion varies between the sequences used.

Suggestions for further testing

- Evaluate the impact of compression on HFR perception, by lowering the 'distribution' bitrate (here 60 Mbit/s) to various levels that can practically be expected.
- Analyse the quality that can be achieved with *p/100 50% shutter* originated material shown on 200 Hz CE displays, as this may make even better use of 100 Hz content production.

7. References

[1] ITU-R BT.2100-2, Image parameter values for high dynamic range television for use in production and international exchange.

https://www.itu.int/rec/R-REC-BT.2100

[2] ITU-R BT.500-13, Methodology for the subjective assessment of the quality of television pictures, January 2012

https://www.itu.int/rec/R-REC-BT.500-13-201201-I/en

Note: All links are current at time of publishing this document.

[9] The 400 m sequence now shows a gain of 0.5 points when comparing the 1080p/50 50% and 1080p/100 100% formats (shown in Figure 9). The overall average for these formats shows 0.6 points gain (illustrated in Figure 10). In both cases the worst-case combination of the 95% confidence intervals was used for the calculation (so being conservative by minimizing the gain).

Annex A: Instructions to participants at the tests

HFR subjective Tests

Duration of about 40 minutes in total

General Remarks

- Welcome to the subjective tests at PTS 2019 / Broadthinking 2019.
- These tests are conducted in the framework of the EBU Video Systems Group.
- The goal is to provide data for the impact of HFR with UHD resolution and HDR.
- We are using a consumer display which can handle HFR with UHD-resolution. All image processing is switched off.
- We are using a very high bit-rate in HEVC to avoid any coding artefacts.
- As the play-out is using transport streams there is no perfect seamless switching between the streams.
- There are some artefacts within the grey sequences, but with no impact on the sequences itself.
- The tests take place according to a standards method defined in the ITU. You will be requested to fill in an electronic form after you have seen different test conditions.
- Please now fill in your name and seat on the device.
- We will start with a short training session to explain how the testing works.

Presentation of sequences and how to use the tablet

- You will see the clips as an A-B-A-B sequence, always with a grey image in between, naming the current clip.
- On the tablet you see 2 horizontal bars named A and B and quality levels ranging from bad to excellent.
- Make your choice comparing A and B. There is a 4 to 5 seconds voting phase at the end of each A-B-A-B.
- When finished, click “submit” and the next comparison is shown.

Training

- All sequences include High Dynamic Range and Wide Colour Gamut. This does not change between sequences.
- HFR provides a better playback of motion. Motion can relate to objects in the picture and to the camera, e.g. when it pans.
- We ask you to score the overall picture quality, focussing on motion.
- Check if objects are blurred, jerky, the motion is natural or disturbed, and so on. You’ll get a feeling for it.
- Check legs, shoes, hurdles, other objects, …
- Don’t be afraid when you spot nothing 😉

Any Questions?

Annex B: Detailed Test Results

January 2019 session - without frame interpolation

	Sequences					
Arrangement	**LongJump**	**Discus**	**Javelin**	**Steeplechase**	**400m**	**Overall**
Ref (2160p/100 100%)	-1.07	2.26	1.19	1.96	-18.04	-2.74
(st.e.m.)	(1.60)	(1.16)	(1.37)	(2.12)	(6.18)	(1.53)
Standard Dev	8.32	6.05	7.11	11.00	32.10	17.72
95% conf	3.14	2.28	2.68	4.15	12.11	2.99
Lower	-4.21	-0.02	-1.50	-2.19	-30.14	-5.73
Upper	2.06	4.54	3.87	6.11	-5.93	0.25
2160p/50 100%	-27.81	-21.22	-29.26	-24.96	-35.56	-27.76
(st.e.m.)	(2.84)	(1.81)	(2.32)	(3.88)	(2.73)	(1.30)
Standard Dev	14.78	9.39	12.04	20.18	14.19	15.12
95% conf	5.58	3.54	4.54	7.61	5.35	2.55
Lower	-33.39	-24.76	-33.80	-32.58	-40.91	-30.31
Upper	-22.24	-17.68	-24.72	-17.35	-30.20	-25.21
2160p/50 50%	-27.67	-18.48	-33.04	-35.15	-37.89	-30.44
(st.e.m.)	(2.95)	(3.84)	(2.95)	(3.61)	(2.91)	(1.56)
Standard Dev	15.33	19.96	15.34	18.76	15.10	18.12
95% conf	5.78	7.53	5.78	7.08	5.70	3.06
Lower	-33.45	-26.01	-38.82	-42.22	-43.59	-33.50
Upper	-21.88	-10.95	-27.25	-28.07	-32.19	-27.39
1080p/100 100%	1.41	3.41	0.48	-9.00	-1.59	-1.06
(st.e.m.)	(2.27)	(3.22)	(2.17)	(2.73)	(1.98)	(1.17)
Standard Dev	11.77	16.74	11.25	14.18	10.29	13.56
95% conf	4.44	6.31	4.24	5.35	3.88	2.29
Lower	-3.03	-2.91	-3.76	-14.35	-5.48	-3.35
Upper	5.85	9.72	4.73	-3.65	2.29	1.23
1080p/50 100%	-33.04	-22.52	-30.96	-33.96	-28.74	-29.84
(st.e.m.)	(2.80)	(2.38)	(2.62)	(3.19)	(3.57)	(1.34)
Standard Dev	14.55	12.39	13.63	16.56	18.54	15.61
95% conf	5.49	4.67	5.14	6.25	6.99	2.63
Lower	-38.52	-27.19	-36.11	-40.21	-35.74	-32.48
Upper	-27.55	-17.85	-25.82	-27.72	-21.75	-27.21
1080p/50 50%	-32.44	-26.33	-34.96	-36.74	-41.37	-34.37
(st.e.m.)	(2.58)	(2.32)	(3.37)	(3.80)	(3.15)	(1.43)
Standard Dev	13.39	12.04	17.50	19.76	16.37	16.58
95% conf	5.05	4.54	6.60	7.45	6.17	2.80
Lower	-37.49	-30.88	-41.57	-44.19	-47.54	-37.17
Upper	-27.39	-21.79	-28.36	-29.29	-35.20	-31.57

March 2019 session - with frame interpolation

	Sequences					
Arrangement	**LongJump**	**Discus**	**Javelin**	**Steeplechase**	**400m**	**Overall**
Ref (2160p/100 100%)	-1.89	0.17	-1.94	5.39	-16.50	-2.96
(st.e.m.)	(2.09)	(1.53)	(2.78)	(2.54)	(8.56)	(2.05)
Standard Dev	8.88	6.48	11.81	10.77	36.30	19.42
95% conf	4.10	2.99	5.46	4.98	16.77	4.01
Lower	-5.99	-2.83	-7.40	0.41	-33.27	-6.97
Upper	2.21	3.16	3.51	10.36	0.27	1.06
2160p/50 100%	-16.06	-13.00	-23.39	-18.39	-33.39	-20.84
(st.e.m.)	(2.52)	(1.33)	(3.01)	(4.87)	(4.10)	(1.67)
Standard Dev	10.71	5.63	12.77	20.68	17.41	15.81
95% conf	4.95	2.60	5.90	9.55	8.04	3.27
Lower	-21.00	-15.60	-29.29	-27.94	-41.43	-24.11
Upper	-11.11	-10.40	-17.49	-8.84	-25.34	-17.58
2160p/50 50%	-9.17	-1.50	-22.78	-22.22	-24.17	-15.97
(st.e.m.)	(3.91)	(2.89)	(4.59)	(4.95)	(4.41)	(2.07)
Standard Dev	16.61	12.24	19.47	21.00	18.72	19.68
95% conf	7.67	5.66	8.99	9.70	8.65	4.07
Lower	-16.84	-7.16	-31.77	-31.93	-32.81	-20.03
Upper	-1.49	4.16	-13.78	-12.52	-15.52	-11.90
1080p/100 100%	-4.56	-2.11	-1.11	-9.17	-6.39	-4.67
(st.e.m.)	(2.12)	(1.99)	(2.77)	(2.29)	(2.41)	(1.07)
Standard Dev	8.99	8.45	11.77	9.73	10.24	10.11
95% conf	4.15	3.90	5.44	4.49	4.73	2.09
Lower	-8.71	-6.01	-6.55	-13.66	-11.12	-6.76
Upper	-0.40	1.79	4.33	-4.67	-1.66	-2.58
1080p/50 100%	-23.17	-13.17	-25.44	-18.50	-28.72	-21.80
(st.e.m.)	(3.31)	(1.98)	(5.37)	(5.02)	(3.09)	(1.83)
Standard Dev	14.05	8.40	22.80	21.29	13.13	17.32
95% conf	6.49	3.88	10.53	9.83	6.06	3.58
Lower	-29.66	-17.05	-35.98	-28.33	-34.79	-25.38
Upper	-16.68	-9.29	-14.91	-8.67	-22.66	-18.22
1080p/50 50%	-20.83	-16.78	-15.89	-27.50	-30.39	-22.28
(st.e.m.)	(5.42)	(2.91)	(3.39)	(3.87)	(4.82)	(1.93)
Standard Dev	23.01	12.33	14.39	16.43	20.45	18.30
95% conf	10.63	5.69	6.65	7.59	9.45	3.78
Lower	-31.46	-22.47	-22.54	-35.09	-39.84	-26.06
Upper	-10.20	-11.08	-9.24	-19.91	-20.94	-18.50