

Deliverable

Deliverable name

D5.3 – Report on QA and Content Preparation Guidelines



| | |
|--------------------------------|--|
| Project Acronym: | IMMERSIFY |
| Grant Agreement number: | 762079 |
| Project Title: | Audiovisual Technologies for Next Generation Immersive Media |

| | |
|--|-------------------------------------|
| Revision: | 2.1 |
| Authors: | Sergio Sanz (SD), Ali Nikrang (AEF) |
| Reviewer: | Mauricio Alvarez-Mesa (SD) |
| Delivery date: | October 2nd 2020 |
| Dissemination level (Public / Confidential) | Public |

Abstract

The Spin Digital real-time 8K HEVC encoder (Spin Enc Live) has been assessed from the subjective quality point of view following the formal recommendation called ITU-T P.913. In the experiments, subjects were asked to assess the subjective quality of a set of 8K video sequences projected on two 8K immersive displays: PSNC 8K video wall and Ars Electronica Deep Space 8K.

Based on the results of the subjective quality assessment, recommendations for the creation of high-resolution immersive content in various target display environments and guidelines for different media content types are provided in the report.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement 762079.

REVISION HISTORY

| Revision | Date | Authors (Entity) | Description of changes |
|----------|-------------|---|--|
| 0.1 | 16 Jun 2019 | Sergio Sanz (SD) | Initial content of the document |
| 0.2 | 06 Mar 2020 | Ali Nikrang (AEF) | First draft |
| 0.3 | 19 Mar 2020 | Mauricio Alvarez-Mesa (SD) | Review |
| 0.4 | 24 Mar 2020 | Sergio Sanz (SD) Ali Nikrang (AEF) | Sections 1-3: final updates after review |
| 0.5 | 27 Mar 2020 | Mauricio Alvarez-Mesa (SD) Sergio Sanz (SD) Ali Nikrang (AEF) | Sections 4-5: final updates after review |
| 1.0 | 30 Mar 2020 | Ali Nikrang (AEF) | Final version |
| 2.0 | 17 Sep 2020 | Sergio Sanz (SD) | Addressing final review comments and suggestions |
| 2.1 | 02 Oct 2020 | Sergio Sanz (SD) | Modified abstract and Introduction |

Contributors

| Contributor's name | Entity | Contact e-mail |
|-----------------------|--------|---|
| Maciej Glowiak | PSNC | mac@man.poznan.pl |
| Eryk Skotarczak | PSNC | eryk@man.poznan.pl |
| Maciej Strozyk | PSNC | mackostr@man.poznan.pl |
| Szymon Malewsky | PSNC | szymonm@man.poznan.pl |
| Sergio Sanz | SD | sergio@spin-digital.com |
| Mauricio Alvarez-Mesa | SD | mauricio@spin-digital.com |
| Christian Södergren | NVAB | christian.sodergren@visualiseringscenter.se |
| Erik Sundén | NVAB | erik.sunden@liu.se |
| Anna Kuthan | AEF | anna.kuthan@ars.electronica.art |
| Ali Nikrang | AEF | Ali.Nikrang@ars.electronica.art |
| Roland Haring | AEF | rolandha@aec.at |
| Johannes Pöll | AEF | johannes.poell@ars.electronica.art |

TABLE OF CONTENTS

| | |
|--|--------------|
| Introduction | 5 |
| Description of Task 5.3 | 5 |
| Relevance of the Task to the Project | 5 |
| Deviations from the Original Plan | 6 |
| Subjective Quality Assessment | 6 |
| Introduction | 6 |
| Motivation and Objectives | 6 |
| Formal Subjective Quality Assessment | 8 |
| Reference Recommendation: ITU-T P.913 | 8 |
| Experimental Viewing Conditions | 8 |
| Test Video Sequences | 12 |
| Encoder Settings | 15 |
| Evaluation Method | 16 |
| Stimuli Randomization | 17 |
| Subjects | 17 |
| Schematic Outline of the Experiment Implementation | 18 |
| Training Session | 18 |
| Official Session | 18 |
| Questionnaire or Interview | 18 |
| Experimental Setup | 18 |
| Subjective Quality Assessment of Spin Enc Live | 19 |
| PSNC Video Wall | 19 |
| Deep Space 8K | 22 |
| Global Results | 24 |
| Color Banding in Images with Gradient Ramps | 27 |
| Introduction | 27 |
| Immersify Logo Animation: Version 1 | 28 |
| Immersify Logo Animation: Version 2 | 30 |
| Encoder Settings for Point-cloud Video | 31 |
| Introduction | 31 |
| Point Size Selection | 31 |
| Encoder Configuration | 32 |
| Encoder Settings for “Immersive Minimalism” | 33 |
| Introduction | 33 |
| Encoder Configuration | 33 |
| Advances over the State of the Art | 34 |
| Status | 35 |
| Expected Results | 35 |
| Deliverable 5.3 | Revision 1.0 |
| | 3/54 |

| | |
|---|-----------|
| Obtained Results | 35 |
| Encoding and Playback Settings | 36 |
| 8K 2D Video | 36 |
| Description | 36 |
| Technical Requirements | 36 |
| Encoder Settings | 37 |
| Playback Settings | 37 |
| 8K 3D (Stereoscopic) Video | 38 |
| Description | 38 |
| Technical Requirements | 38 |
| Encoder Settings | 39 |
| Playback Settings: Minimum Requirements | 39 |
| Playback Settings: Immersive Environments | 39 |
| High-resolution 360° Playback | 40 |
| Description | 40 |
| Technical Requirements | 40 |
| Encoder Settings | 41 |
| Playback Settings | 41 |
| Point-cloud Video Playback | 42 |
| Description | 42 |
| Technical Requirements | 42 |
| Encoder Settings | 43 |
| Playback Settings | 43 |
| Interactive Video Playback | 44 |
| Description | 44 |
| Technical Requirements | 44 |
| Encoder Settings | 45 |
| Playback Settings | 45 |
| Status | 46 |
| Expected Results | 46 |
| Obtained Results | 46 |
| Content Preparation Guidelines | 47 |
| Point Cloud | 47 |
| 8K 2D/3D Filming | 47 |
| Ambisonic Sound Production | 48 |
| Dome and Interactive Content | 49 |
| Interactive Authoring, Deep Space DevKit | 50 |
| AI-Based Super Resolution Upscaling | 51 |
| Status | 53 |
| Expected Results | 53 |
| Obtained Results | 53 |

1 Introduction

In this document the progress made in Task 5.3 on “Quality Assessment and Content Preparation Guidelines” will be reported. It includes activities that have been performed within the context of the task, their results and their deviations from the original plan. It also includes a set of guidelines for content preparation for media artists. Based on the results of subject quality assessment tests of the Spin Digital real-time 8K encoder reported in Section 3, some recommendations for encoding and playback immersive content is given in Section 4 based on the use cases targeted for the project. Finally, best practises and recommendations for creating high resolution video content for media artists will be described in Section 5.

2 Description of Task 5.3

According to the Grant Agreement (GA), the objective of Task 5.3 on “Quality Assessment and Content Preparation Guidelines” is:

“In this task the new content produced in Tasks 5.1 and 5.2 will be encoded with the enhanced HEVC encoder developed in WP4. Specific quality-rate points will be selected according to the demands of end-users and the system requirements for the envisioned demonstrations (Tasks 5.4 and 5.5). Optimal configurations of the encoder will be found for maximizing the subjective quality given a target bitrate, while at the same time guaranteeing real-time operation of the decoder. An informal subjective test will be prepared on the target display environments in order to validate the improvements implemented in the encoder in terms of subjective quality.

Encoding guidelines and workflow optimizations will be prepared to assist content creators for preparing the immersive content for delivery, and content exhibitors to adjust the content for maximum quality and real-time playback.”

And the expected outcome is:

“Guidelines for encoding and decoding immersive content, as well as media production workflows experiences. As a best practice reference, the information will also be published on the project website.”

2.1 Relevance of the Task to the Project

Objective 1 of the project is defined as “improving the quality of immersive media using advanced compression technology” which is directly related to the topic of this deliverable. In this task we use state of the art methods to estimate the quality of the results obtained with different compression methods applied in this project. The progress of quality assessments includes subjective tests and evaluation methods that estimate the perceived quality of a video by human observers. Objective

assessments have been performed as well and are presented in D4.3 - Report on Live Streaming Server Integration.

2.2 Deviations from the Original Plan

D5.3 was originally planned for the period of Month 18 to Month 23 of the project. The consortium agreed to an extension of three months (until June 2020) for additional dissemination activities as it is described in amendment 2. This includes a corresponding extension of task 3.5 and its deliverables until March 2020.

3 Subjective Quality Assessment

3.1 Introduction

Deliverable *D4.3 on "Report on Live Streaming Server Integration"* describes the work performed in *Task 4.3 - Real-time Encoding* with regard to the assessment of 8K HEVC real-time encoders in terms of encoding speed and compression efficiency. This latter criterion is measured by means of the so-called BD-rate metric or, in other words, the average bitrate increase that a test encoder produces with respect to a baseline encoder for the same *objective quality* which is typically measured using Peak Signal-to-Noise Ratio (PSNR).

Quality assessment of video encoders can also be conducted through subjective experiments with the objective of estimating the so-called Mean Opinion Score (MOS). In the context of our work, this metric reflects the average opinion given by a set of subjects concerning the quality perceived of a video that is encoded at a certain bitrate. Although the MOS is more correlated to human judgments compared to distortion metrics (PSNR, MSE, SSIM¹), conducting subjective testing requires much more logistic effort and time for collecting a representative number of test subjects and preparing test sessions.

In the following subsections we will describe all the work performed in Task 5.3 on subjective quality assessment of the encoding solutions that have been developed in WP4 and used in some dissemination activities.

3.2 Motivation and Objectives

According to the description of Task 5.3 in the GA, an informal subjective test shall be conducted on the target display environments, in order to validate the enhancements implemented in the 8K real-time encoder developed in WP4 (aka **Spin Enc Live**). Another objective of this task on subjective quality is to find a good encoder configuration (mainly bitrate) that can guarantee high subjective quality for the target use cases envisioned in the project

As target display environments, we chose the PSNC 8K video wall and the AEF Deep Space 8K and, as test video sequences, representative immersive content in 8K resolution produced by state-of-the-art video acquisition technologies, such as 8K camera, CGI, laser scanner, and timelapse.

¹ Z. Wang, *Image Quality Assessment: From Error Visibility to Structural Similarity*, IEEE Transactions on Image Processing, vol 13, no. 4, 600-612, 2004

The most common approach to assess the subjective quality produced by an encoder is by means of the MOS score computed at different bitrates. Subjective tests can be performed in a formal or an informal way. As already mentioned in Section 3.1, formal subjective tests produce more accurate results but these are very time consuming. Alternatively, informal subject tests reduce time but at the expense of less accurate results. Although in the GA we anticipated an informal test for conducting subjective quality assessments, after an internal discussion we finally decided to carry out a formal subjective test for the sake of accuracy in the results.

The formal subjective test performed in Task 5.3 aimed at assessing the quality produced by Spin Enc Live and finding out a good bitrate configuration for live applications. In addition, after some internal discussions between SD, PSNC, and AEF, we identified other cases whose visual effect should be assessed from a subjective point of view. Those cases are: 1) color banding effect in pictures with gradient ramps, 2) encoding configuration for special content such as point-cloud video renderings, and 3) encoding configuration for Theresa Schubert’s content titled “Immersive Minimalism”. In this way, we have divided the whole subjective quality assessment task into five different experiments that are listed in the following table. The partners involved in the tests as well as the type of experiment (formal or informal) are also specified.

Table 1. Set of subjective tests conducted in Task 5.3

| Title | Partners involved | Type | Description |
|--|-------------------|----------|--|
| Subjective quality assessment of Spin Enc Live | SD, PSNC, AEF | Formal | Assess the quality of the Spin Digital real-time encoder from a subjective point of view on target display environments (PSNC video wall and AEF Deep Space 8K) using immersive content. |
| Color banding effect in images with gradient ramps | SD, AEF | Informal | Pictures created with gradient ramps are very sensitive to color banding after encoding, so the aim is to analysis how this artefact affects the perceived quality and proposed mechanisms to reduce such an effect |
| Encoder settings for point-cloud video | SD, PSNC | Informal | Point cloud from laser scanning render to video is a special content that is very challenging for the HEVC encoder. The aim is to specify the best configuration in terms of point size and encoder settings to properly encode this type of content |
| Encoder settings for Immersive Minimalism | SD, PSNC, AEF | Informal | The Thesa Schubert’s content titled “Immersive Minimalism” is also very challenging from the encoding perspective. The aim is to specify the best configuration in Spin Enc to properly encode this content |

Before analysing in detail the results of the subjective tests outlined in the table, the methodology that we followed for conducting formal subjective tests and the decisions we made to adapt this methodology to the scope of Task 5.3 are described in the next subsection.

3.2.1 Formal Subjective Quality Assessment

3.2.1.1 Reference Recommendation: ITU-T P.913

Recommendations ITU-T BT.500² and ITU-T P.910³ have been used for many years to perform subjective quality assessment of video codecs. These recommendations specify exact viewing conditions (viewing distance, screen luminance, room illumination) in quiet and non-distracting environments in order to reduce the influence of the environment from the experiment. However, our target environments, as public spaces they are, may have their own viewing conditions for highly immersive experience. This is, for example, the case of Deep Space 8K, where the recommended viewing distance can be shorter than what BT.500 or P.910 recommend.

Recommendation ITU-T P.913 attempts to solve this issue, since it accepts both public environments and controlled environments, such as pristine laboratories or simulated rooms. The subjective test also addresses the scope described in this recommendation:

- *“The Recommendation describes methods to be used for subjective assessment of the audiovisual quality of Internet video and distribution quality”*. Although the videos to be used in our tests will not be transmitted or streamed but reproduced locally on a PC, their qualities (or bitrates) are considered common for 8K live Internet and distribution-quality 8K TV.
- *“The Recommendation may be used to compare audiovisual device performance in multiple environments and to compare the quality impact of multiple audiovisual devices”*. The audiovisual device will be the 8K real-time software encoder (Spin Enc Live) and its performances in terms of compression efficiency will be assessed in two environments (PSNC 8K video wall, Deep Space 8K).
- *“It is appropriate for a wide variety of display technologies, including flat screen, 2D, 3D, multi-view and autostereoscopic”*. Array of projectors for 8Kp60 video playback will be used in the target environments.

In addition, Rec. P.913 makes reference to other recommendations, especially BT.500 and P.910. Among other topics, this recommendation also accepts the evaluation methods and rating scales described in BT.500 and proposes the video selection methodology provided in P.910.

3.2.1.2 Experimental Viewing Conditions

In Table 2 and Figures 1 and 2, the experimental viewing conditions are described. The two immersive spaces under consideration are controlled environments, because of the following:

- The illumination in both rooms was non distracting and dark (conditions also recommended in BT.500 and P.910);
- The rooms were comfortable and quiet and had no distracting elements on the walls;
- Only the participants and the experimenter were allowed to be in the room.

As can be observed in Figure 1, in the PSNC 8K video wall room there were also some low directional

² [Rec. ITU-R BT.500-14, Methodology for the Subjective Assessment of the Quality of Television Pictures, 2019.](#)

³ [Rec. ITU-T P.910, Subjective Video Quality Assessment Methods for Multimedia Applications, 2008.](#)

spot lights pointing at the table for the participants to easily annotate scores on the paper ballot.

Table 2. Summary of the experimental viewing conditions

| | PSNC 8K video wall | AEF Deep Space 8K |
|---------------------|--|--|
| Type of environment | Controlled | Controlled |
| Room conditions | - Non-distracting environment - Totally dark with a few low directional spot lights pointing at the table for participants to annotate scores on the ballot | - Non-distracting environment - Totally dark |
| Projection system | - 12 rear projectors BARCO F50 - 2560x1600 pixels each - Measured luminance: 40 cd/m ² | - 4 front projectors Christie Mirage 304K - 4096x2160 pixels each - 15.000 ANSI lumen each |
| Noise level | Quiet | Quiet |
| Screen size | Width = 6.0 m Height = 2.8 m Diagonal = 6.6 m | Width = 16.0 m Height = 9.0 m Diagonal = 18.4 m |
| Screen resolution | 7680x4320 pixels | 6467x3830 pixels (only wall projection) |
| Viewing distance | 4.2 m (1.5 x H) | 8.75 m (\approx 1.0 x H) |
| Field of view | 64.0° to 71.1° | 82.3° to 84.9° |
| Viewing positions | 6 viewing points | 7-8 viewing points |

The projection system of the PSNC video wall consists of 12 rear projectors BARCO F50 forming a layout of 3 rows of 4 projectors. The size of the video wall is 6 m x 2.8 m with a total resolution of 8192x4320 pixels after blending. As shown in Figure 1, six participants were positioned along a 3.8-meter viewing line. The viewing distance (VD) was set to 4.2 m (1.5 x H) and the resulting field of views (FoV) for a 6-meter-width screen ranged from 64.0°, at the left and right extremes of the viewing line, and 71.1°, at the center of the viewing line. This distance of 4.2 m was carefully chosen to provide a good trade-off between cinematic view (40°-45°⁴), which allows to see the whole screen without moving the head, and immersive effect (96°-100° for 8K TV in 16:9^{5,6}). As can be seen in Table 3, the viewing distance recommended by PSNC complies with the recommendations given in P.913, P.910, and BT.500. According to P.913, a minimum viewing distance shall correspond to the Least Distance of Distinct Vision (LDDV), which is the closest distance someone with normal vision (or

⁴ [F. Friedrich, 8K Resolution: Hype of Benefit, Nov. 15th 2019](#)

⁵ T. Yamashita, K. Masaoka, K. Ohmura, M. Emoto, Y. Nishida, M. Sugawara, *Super hi-vision - video parameters for next-generation television*. SMPTE Motion Imaging J., 121 (4) (2012), 63–68.

⁶ [Report ITU-R BT.2246-6, The present state of ultra-high definition television, 2017](#)

Snellen 20/20 vision⁷) can look at something comfortably. In the case of 8K TV, it corresponds theoretically to a FoV of 120°, but in practice 96° (0.75 x H) is recommended.

Table3. Viewing distances recommended by reference recommendations and compliance with the ones set for the subjective test

| Recommended | | PSNC 8K video wall | Deep Space 8K |
|-------------|--|--------------------|-------------------|
| P.913 | > LDDV (0.75 x H for 8K TV) | 4.2 m (1.5 H) | 8.75 m (≈1.0 x H) |
| P.910 | 1.0-8.0 x H | | |
| BT.500 | 1.5-2.0 x H (for large screen imagery) | | |

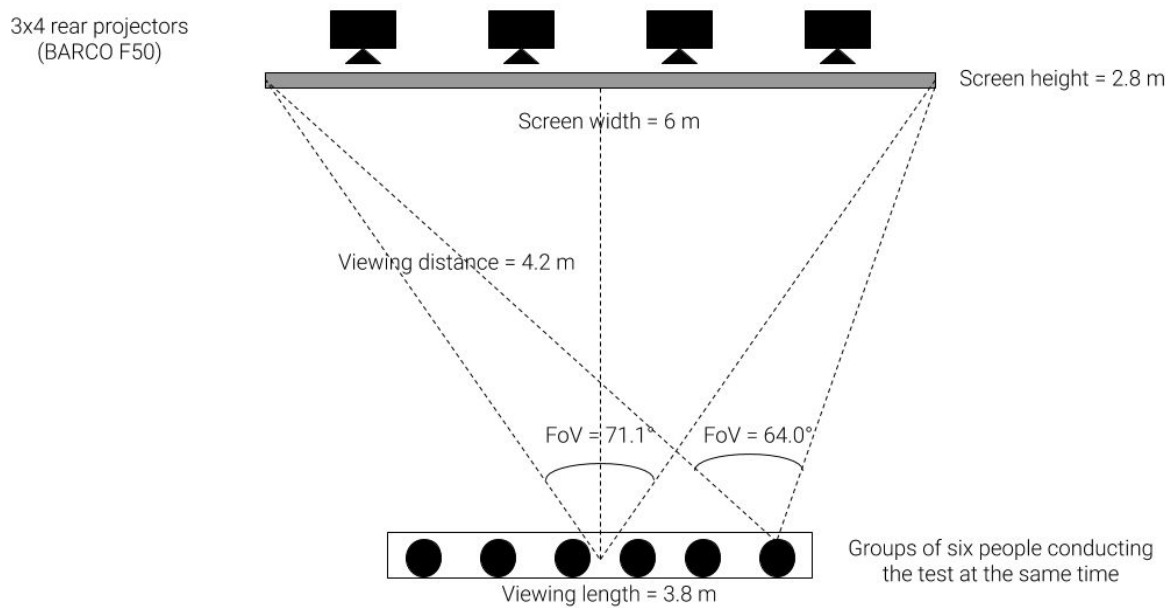


Figure 1. Viewing conditions in the PSNC video wall environment

⁷ S. T. McCarthy, *How Independent are HDR, WCG, and HFR in Human Visual Perception and the Creative Process?*, Arris, 2016

The wall projection system in Deep Space includes four front projectors Christie Mirage 304K as shown in Figure 2. These are located 9 m away from the screen at a height of 4.78 m. The size of the screen is 16 m x 9 m with a total resolution of 6467x3830 pixels after blending. Up to 8 participants performed the test at the same time. Those people sat 8.75 meters away from the screen along a viewing line of around 5 meters long. For a 16-meter-width screen size the resulting FoVs ranged from 82.3° to 84.9° which correspond to the extreme sides and the center of the viewing line, respectively. The viewing distance of 8.75 meters ($\approx 1.0 \times H$) is considered by the AEF team as the sweet spot for the Deep Space 8K environment to achieve an immersive experience. According to Table 3, this distance is within the range recommended in P.913 and P.910.

Before sending the decoded picture to the screen, the video render performs a downscaling operation in order to accommodate the video resolution (7680x4320 pixels) to the total resolution of the wall screen after blending (6467x3830 pixels).

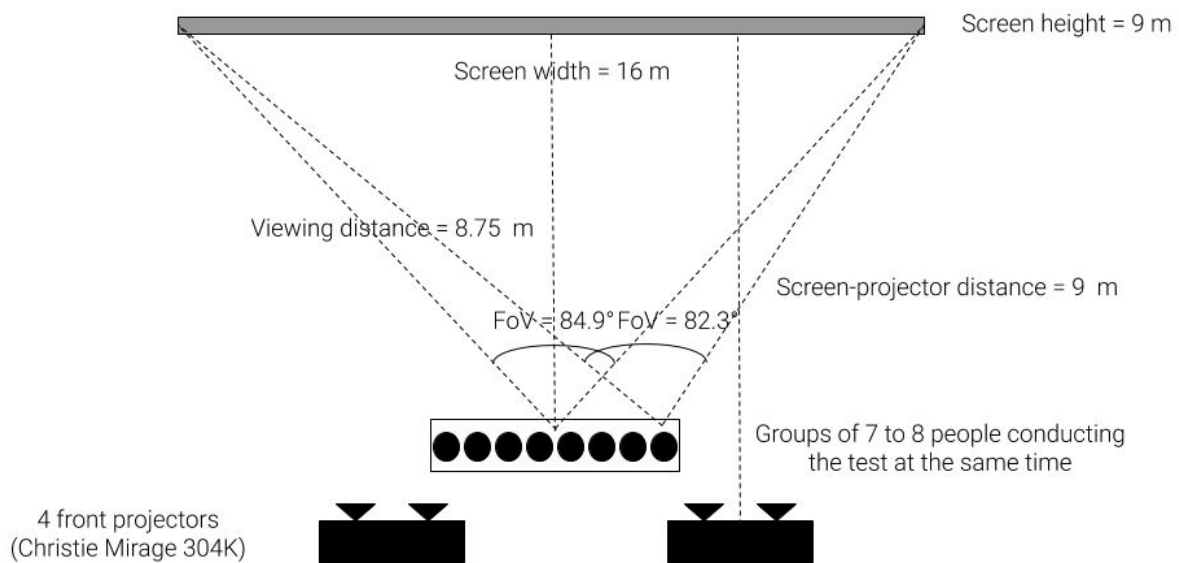





Figure 2. Viewing conditions in the Deep Space 8K environment

3.2.1.3 Test Video Sequences

As shown in Table 4, a total of six 8Kp60 4:2:0 10-bit video sequences were used for the experiments. An additional one was also used for the training session. Each sequence had a duration of 10 seconds, as recommended in P.913.

The sequences were created using different acquisition and production technologies, in particular: camera footage (Follow Car, MC2, Follow Car 2), timelapse (Berlin and Island in the Sky II), point cloud from laser scanning rendered to video (Cathedral), and CGI (Singing Sand).

Table 4. Set of test video sequences

| | |
|--|--|
| <p>Berlin (timelapse)</p> <p>Official</p> |  |
|--|--|

Follow Car (footage)

<https://youtu.be/SCw9i3T8Ff4>

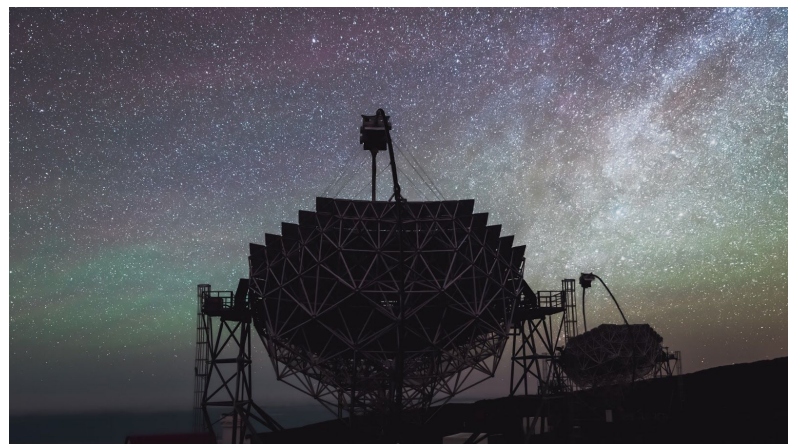
Official



Island in the Sky II (timelapse)

<https://youtu.be/Ci8P7gMJbVg>

Official


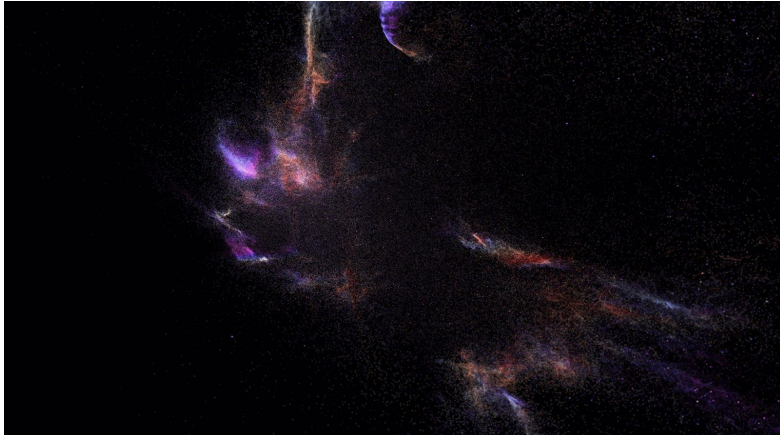



MC2 (footage)

<https://youtu.be/rCeau4OrH3w>

Official



| | |
|---|--|
| <p>Cathedral (point cloud) https://youtu.be/hcoFpMZSJQ4</p> <p>Official</p> |  |
| <p>Singing Sand (CGI) https://vimeo.com/347031563</p> <p>Official</p> |  |
| <p>Follow Car (footage) https://youtu.be/SCw9i3T8Ff4</p> <p>Training</p> |  |

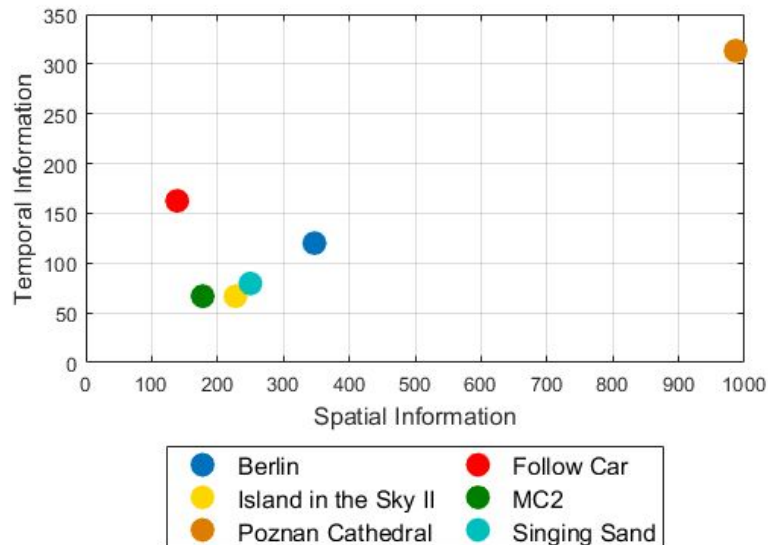


Figure 3. Spatial and temporal perceptual information of the six 8Kp60 video sequences used in the official test

Based on Rec. P.910, the spatial and temporal perceptual information (SI and TI) of these sequences were computed and graphically represented in Figure 3. High values of SI and TI stand for high spatial and temporal complexities, respectively. Note that the SI and TI values of Cathedral are much higher than the rest. The reason behind this is because the content is made of tiny point clouds from a laser scanner rendered to video. As a result, the textures present very high-frequency components with a distribution of points that changes a lot over time. This type of content is in fact very challenging for video encoders based on H.264, HEVC, or AV1, since the supported coding tools do not work efficiently.

In the case of Singing Sand, the SI and TI values are not so high, despite the fact that this content is also made of points. The difference with respect to Cathedral is that the particles in Singing Sand are more separated in space and the background areas are uniform and big.

As recommended in P.913, a training session was also carried out for the users to get familiar with the testing methodology and voting procedure. For this training session a sequence different to those used for the official test was used, in particular “Follow Car 2”, the last one listed in Table 4. Its SI and TI values are, respectively, 119.3 and 91.5.

The original duration of each sequence was 17 seconds, but only the 10 seconds in the middle were used. After encoding the videos at different bitrates (see next section), the first 3-4 seconds and the last 3-4 seconds were removed from the content in order to avoid potential quality issues due to bitrate fluctuations at the beginning and the end of the encoding process.

3.2.1.4 Encoder Settings

For the objectives of the subjective test, the encoding bitrates should cover a range typically used for live streaming and broadcasting of 8K video. Assuming as a reference point 85 Mbps, the bitrate

recommended by NHK for 8K broadcasting⁸, we chose some bitrates around this value, in particular: 25, 40, 60, 85 (reference), 100, 150 Mbps. In order to verify the suitability of the proposed set of bitrates, a pre-pilot test was conducted by a video coding expert at Spin Digital using a 2x2 4Kp60 monitor with dimensions 1.4 x 0.8 m at a viewing distance of 1 m. This test confirmed the presence of noticeable coding artefacts in some sequences encoded with Spin Enc Live at 25 Mbps and no visible errors at high bitrates in any of the tested sequences.

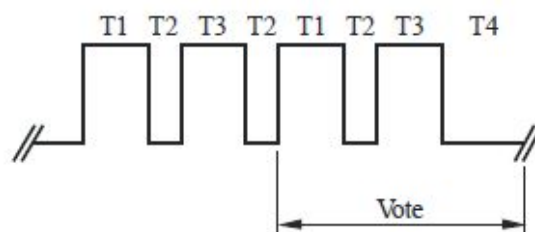
The following table summarizes the encoding parameters we finally decided for the formal subjective test. These parameters correspond to the ***broadcast streaming scenario***.

Table 5. Experimental setup for the formal subjective quality assessment

| | |
|-----------------------------|--|
| Video encoder | Spin Enc Live - normal |
| Video coding scheme | HEVC Main 10 Profile (4:2:0 10-bit) |
| Rate control | CBR |
| GoP structure | 16 frames (hierarchical) |
| Intra period | 1 second |
| Bitrates | 6 bitrates: 25, 40, 60, 85, 110, 150 Mbps |
| Test sequences | 6 videos: 8Kp60, 10 seconds (600 frames) |
| Total number of evaluations | 36 evaluations (1 encoder x 6 bitrates x 6 videos) |

3.2.1.5 Evaluation Method

An evaluation method based on the *Double Stimulus Impairment Scale (DSIS) - Variant II* -, described in BT.500, was used. Also known as *Degradation Category Rating (DCR)* in P.913, this method is suggested if video sequences are under test.



⁸ Y. Sujito, S. Iwasaky, K. Chida, K. Iguchi, K. Kanda, X. Lei, H. Miyoshi, K. Kazu, *Video Bit-rate Requirements for 8K 120-Hz HEVC/H.265 Temporal Scalable Coding: Experimental Study based on 8K Subjective Evaluations*, APSIPA Transactions on Signal and Information Processing, 2019.

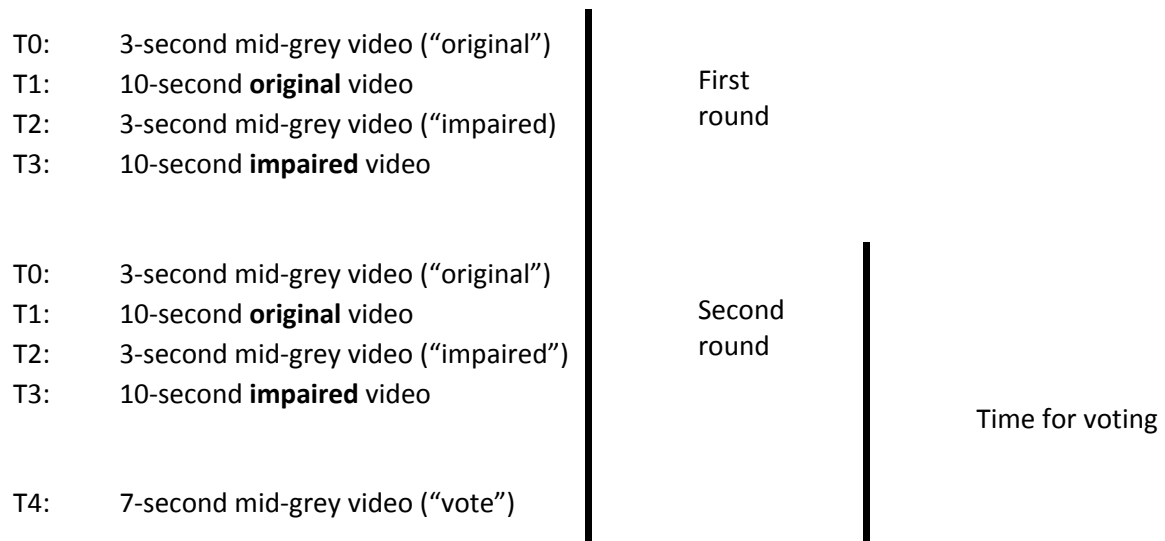


Figure 4. DSIS method - Variant II (source BT.500)

As shown in Figure 4, this evaluation method consisted of a two-round sequence of clips. In each round, the original video was shown first (T1) and then the impaired version (T3). Between the original and impaired videos, 3-second mid-grey clips with messages “next video: original” (T0) and “next video: impaired” (T2) were shown. After the second round, a 7-second mid-grey clip with message “Vote” (T4) was shown. Those messages shown in T0, T2, and T4, although not specified in BT.500, turned out to be very useful when the users started to have fatigue after several evaluations.

People were allowed to vote from the beginning of the second round and what they rated was the impairment of the second stimulus (T3) in relation to the reference one (T1) by using the following five-grade scale:

- 5 “imperceptible”
- 4 “perceptible but not annoying”
- 3 “slightly annoying”
- 2 “annoying”
- 1 “very annoying”

This method was repeated 36 times. During the training session conducted before the official evaluations, the users were instructed on this evaluation method.

3.2.1.6 Stimuli Randomization

Following the recommendation given in P.913 for reducing the impact of playback ordering effects, two random orderings of the video sequences (stimuli) were created: R1 and R2. Each ordering was divided into two groups of 18 videos each: group A and B. In this way, four different orderings were generated and named as:

- Session 1: R1A (first half of videos) + R1B (second half of videos)
- Session 2: R1B + R1A
- Session 3: R2A + R2B
- Session 4: R2B + R2A

For each environment under assessment, the set of participants was divided into groups of 6 to 8

participants (see Figures 1 and 2). People belonging to a group assessed simultaneously the quality of the videos of a session different to that used for the previous group, and so on. The next figure illustrates the way we organized people and video clips to prevent bias on playback ordering.

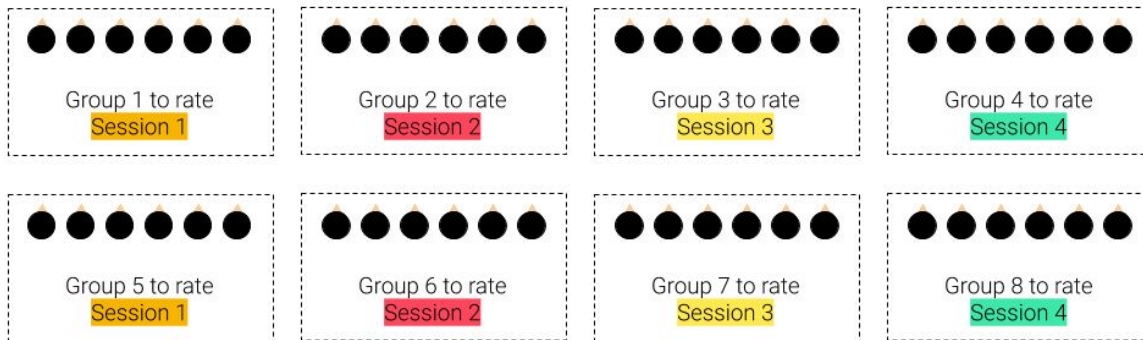


Figure 5. Stimuli randomization and organization of subjects into groups

3.2.1.7 Subjects

According to Rec. P.913, at least 24 subjects in a controlled environment. As the PSNC 8K video wall and the AEF Deep Space 8K are addressed not only to the general public but also to professionals in the media industry, docents, and researchers, two types of subjects were considered: non experts and experts. Ideally, video coding experts should participate in the evaluation, since the main target of the test is to assess the quality of video encoder⁹. The expert people that could be collected were professionals in the media industry and, in particular, in production of immersive content in 8K (and beyond) using state-of-the-art capturing technologies, such as 8K cameras, laser scanners, and Unity/Unreal CGI.

Rec. P.913 also specifies that the set of participants should be well balanced in age distribution and gender.

3.2.1.8 Schematic Outline of the Experiment Implementation

Each group-of-subject's participation in a session consisted of the following stages:

1. Informed consent and instructions
2. Training session
3. Official session
 - Vote Session X - Video 1
 - Vote Session X - Video 2
 - ...
 - Vote Session X - Video 36
4. Questionnaire or interview

3.2.1.9 Training Session

As already mentioned in Section 3.2.1.3, a training session was conducted for the participants to get

⁹ Y. Sujito, S. Iwasaky, K. Chida, K. Iguchi, K. Kanda, X. Lei, H. Miyoshi, K. Kazu, *Video Bit-rate Requirements for 8K 120-Hz HEVC/H.265 Temporal Scalable Coding: Experimental Study based on 8K Subjective Evaluations*, APSIPA Transactions on Signal and Information Processing, 2019.

familiar with the DSIS - Variant II method (Figure 4) and voting procedure. Follow Car II was the sequence that we finally selected for this session. Three versions of this video were generated at different bitrates: 25 Mbps (lowest quality), 60 Mbps (middle quality), 150 Mbps (highest quality).

3.2.1.10 Official Session

By means of the DSIS - Variant II method, the quality of each video under assessment referred to the original version was rated from 1 (“very annoying”) to 5 (“imperceptible”) and annotated on a paper ballot. Besides, the paper ballot included some fields to fill in the age, gender, and level of expertise for statistical purposes.

The sessions were expected to last between 50 and 60 minutes. Breaks were also allowed during a session in case that someone felt tired after several evaluations.

3.2.1.11 Questionnaire or Interview

Although not mandatory in P.913, people who performed the subjective test could annotate on the paper ballot their feedback or directly tell their impressions about the test to the experimenters. The feedback given by the users will be reported in the section that describes and analyses the results.

3.2.1.12 Experimental Setup

One of the main drawbacks of relying on the DSIS method for subjective quality assessment is that the original content should be presented ideally in raw YUV format, not in compressed HEVC. This requirement led us to use the Spin Digital raw player (*Spin Render*), which has been optimized for rendering uncompressed video up to 8Kp60 4:2:0 10-bit. The disk transfer data rates for 8Kp60 4:2:0 10-bit video are up to 5972 MB/s (depending on the pixel format). Although it is possible to achieve those data rates using state-of-the-art SSD arrays, we did not have those resources available in our consortium. As an alternative solution we used a light compression method based on the BC4 format for texture compression which has been evaluated as visually lossless and that can result in compression ratios of up to 4:1 (depending on the pixel format). The data rate of 8Kp60 4:2:0 10-bit video using BC4 is 1400 MB/s which can be handled with a single SSD disk. A quality analysis of the BC4 format was presented in D3.1 - Report on Decoder and Video Rendering for VR - Part 1.

The hardware and software platforms used for the subjective test in the environments at PSNC and AEF are given in Table 6.

Table 6. Software and hardware requirements

| | PSNC 8K video wall | Deep Space 8K |
|--------------|----------------------------------|--------------------------|
| Video player | Spin Render v2.2-beta-2-gfce25c1 | |
| OS | Windows 10 Pro (v.1903) | Windows 10 Pro (v. 1909) |
| CPU | 2 x Intel Xeon Platinum 8168 | 2 x Intel Xeon Gold 6140 |
| GPU | NVIDIA Quadro P5000 | 2 x NVIDIA Quadro P6000 |
| Memory | 192GB | 96GB |

| | | |
|----------|-------------------------------------|--|
| SSD Disk | 2 x Intel SSD DC P3600 Series 800GB | 3 x Intel SSD DC P4500 1TB MegaRAID 9460-16i (Raid 5) |
|----------|-------------------------------------|--|

3.2.2 Subjective Quality Assessment of Spin Enc Live

3.2.2.1 PSNC Video Wall

The assessment of 36 videos lasted approximately one hour. A total of 34 subjects conducted the test. If we classify the participants by gender, 27 of them were men and 7 women, and, by the level of expertise, 27 were non experts and 7 experts. The age distribution is depicted in Table 7.

Table 7. Age Distribution - PSNC 8K video wall

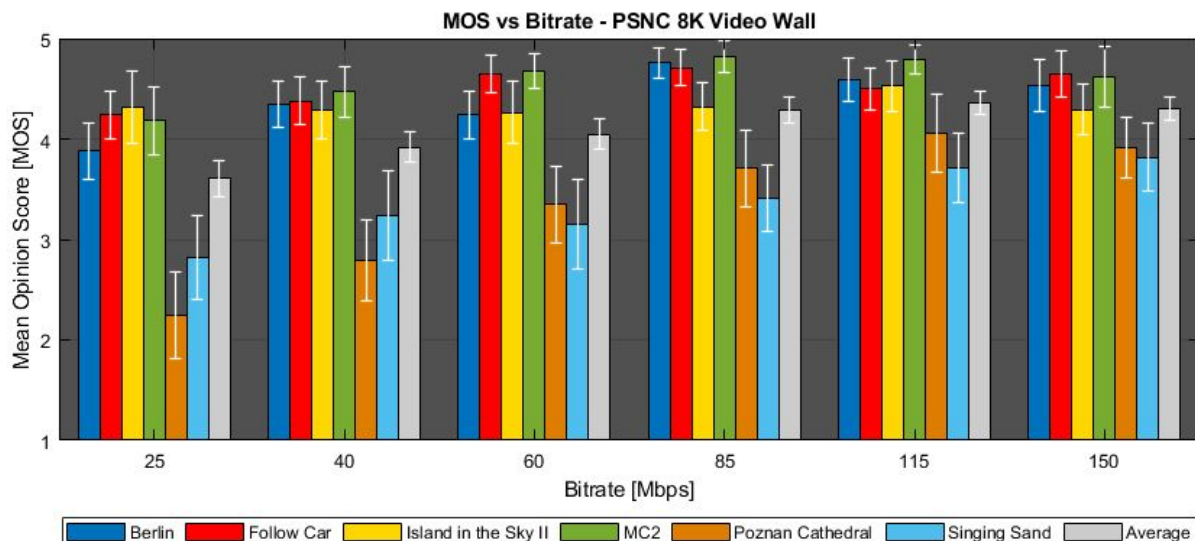
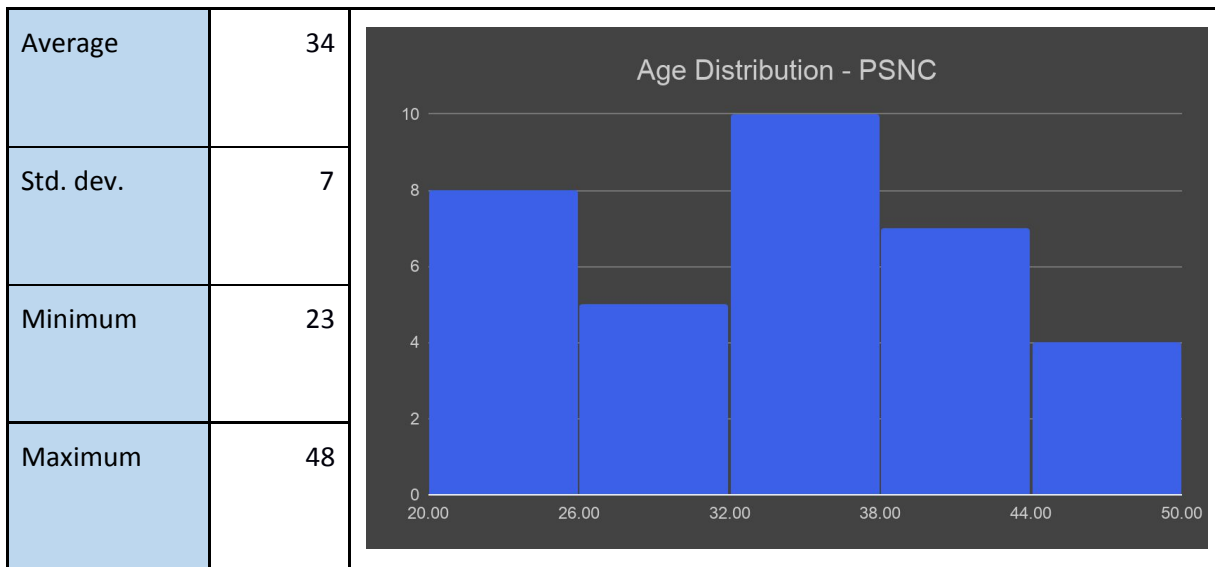


Figure 6. MOS versus Bitrate - PSNC 8K video wall

The results in terms of Mean Opinion Score (MOS) versus bitrate are illustrated in Figure 6. As we can observe, the MOS values were in general very high for video footage (Follow Car and MC2) and timelapse (Berlin and Island in the Sky II). Note that the results for Berlin are fluctuating with the

bitrate. It is not clear yet the reason behind this behavior, but this proves that evaluating timelapse content is difficult from the subjective quality perspective.

The MOS values for Cathedral and Singing Sand are, in contrast, lower than those for the rest of videos under assessment. Since these two sequences present highly discontinuous textures, the resulting PSNR produced by the encoder at a given bitrate drops considerably. It is also important to highlight that Singing Sand was in general underscored. Some of the users disliked this abstract content, because they could not understand what it was about. Perhaps a duration longer than 10 seconds would have been required for Singing Sand.

Although some sequences showed lower MOS values at the highest bitrate than those at lower bitrate, in most cases these differences are negligible because the MOS values exceed 4.5, the threshold of visibility¹⁰, which means that beyond 4.5 coding impairments are barely visible.

According to Figure 6, we can also conclude that the bitrate saturation point from which the average MOS hardly increases is 85 Mbps. At this bitrate most of the sequences generated a MOS higher than 3.5.

Table 8. 95% Confidence Interval - PSNC 8K video wall

| | 25 Mbps | 40 Mbps | 60 Mbps | 85 Mbps | 110 Mbps | 150 Mbps |
|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| Berlin | 3.88±0.28 | 4.35±0.23 | 4.24±0.24 | 4.76±0.15 | 4.59±0.21 | 4.53±0.26 |
| Follow | 4.24±0.24 | 4.38±0.24 | 4.65±0.19 | 4.71±0.18 | 4.50±0.21 | 4.65±0.23 |
| Island | 4.32±0.36 | 4.29±0.29 | 4.26±0.31 | 4.32±0.24 | 4.53±0.25 | 4.29±0.25 |
| MC2 | 4.18±0.34 | 4.47±0.25 | 4.68±0.17 | 4.82±0.16 | 4.79±0.14 | 4.62±0.30 |
| Cathedral | 2.24±0.43 | 2.79±0.40 | 3.35±0.38 | 3.71±0.38 | 4.06±0.39 | 3.91±0.30 |
| Singing | 2.82±0.42 | 3.24±0.45 | 3.15±0.45 | 3.41±0.33 | 3.71±0.35 | 3.82±0.34 |
| Average | 3.61±0.18 | 3.92±0.15 | 4.05±0.15 | 4.29±0.13 | 4.36±0.12 | 4.30±0.12 |

Figure 6 and Table 8 show the 95% confidence interval (CI) based on the Student's t-distribution for each video sequence and bitrate. The narrower the CI is the more precise the estimation of the MOS is. As we can see, for camera footage and timelapse the error margins around the average are in most cases below 0.3 and, for Cathedral and Singing Sand the errors get bigger until ±0.38 to ±0.45. We can also observe that in general the CIs are wider at 25 Mbps than at the rest of bitrate and the narrowest points are achieved at 85 Mbps.

A summary of the feedback given by the subjects is listed below:

1. Many people complained about jumps in the Singing Sand excerpt. Without music it looks like an error. They said it was annoying in the original and hard to assess because of it. The experimenter guesses that, as they did not know the nature of impairment, assuming also some temporal issues, they tried to see if in the impaired version it is even more annoying or

¹⁰ ITU-R BT.500-14, *Methodologies for the subjective assessment of the quality of television images*, October 2019: <https://www.itu.int/rec/R-REC-BT.500>

not.

2. Similar comments were given to the Island in the sky excerpt, but to much less extent.
3. The general opinion was that in natural images it was very hard to see the differences and that the notes are very uncertain. In contrast the computer generated images were considered much easier to assess.
4. Many of the participants admitted that with time they started to see more. That they selected some spots to observe, where they could see the differences in quality (clouds, fence, etc.).
5. In Singing Sand there were some artefacts near the edges in the black background, quite easy to overlook, but once you see them - quite annoying and easy to spot in next voting.

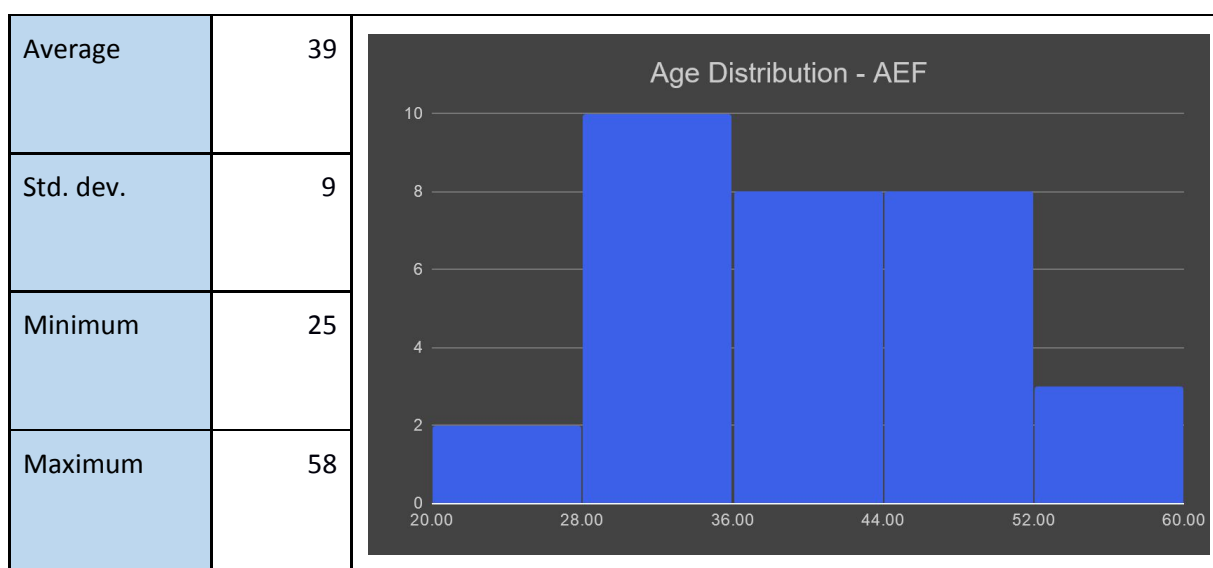
From all these comments it is important to highlight the following conclusions:

1. **Singing Sand caused rejection** but errors were easy to score once identified them
2. It was **difficult to spot errors in natural content**
3. The participants **could detect more and more errors with the time** after several repetitions of the same sequence. Fortunately, the stimuli randomization strategy described in Section 3.2.1.6 helped to reduce this bias.

3.2.2.2 Deep Space 8K

The subjective evaluation of 36 sequences also lasted one hour in the Deep Space 8K. A total of 31 subjects conducted the test. Approximately one third of the subjects were women (11) and the rest men (20). If we divide them into expertise levels, 22 were non-experts and 9 experts. The age distribution is depicted in Table 9.

Table 9. Age Distribution - Deep Space 8K



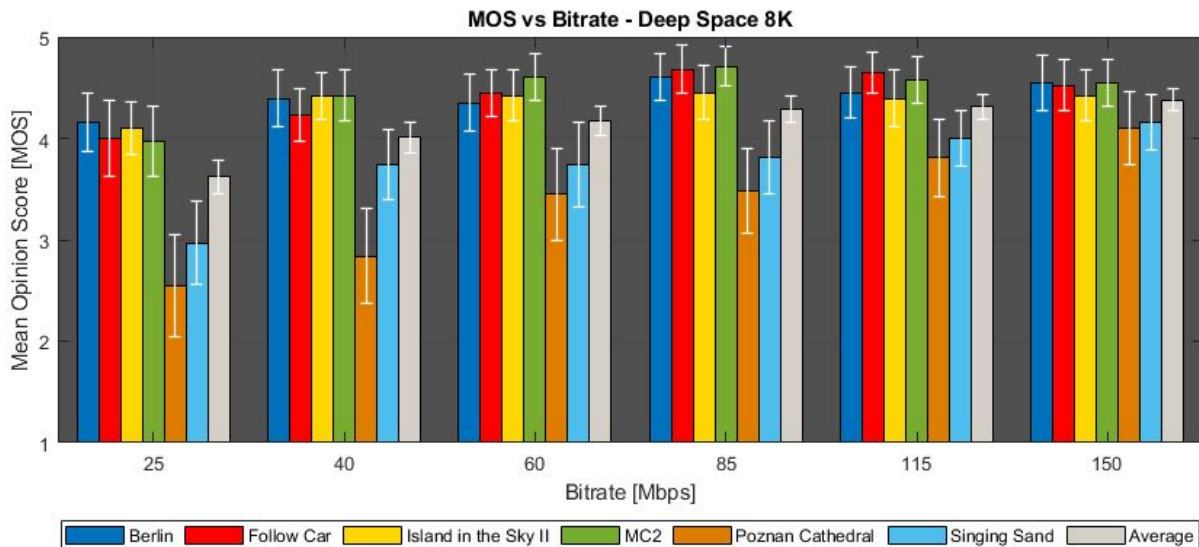


Figure 7. MOS versus Bitrate - Deep Space 8K

The MOS-vs-bitrate results corresponding to the Deep Space 8K environment are depicted in Figure 7. As can be shown, the sequences achieved MOS scores very close to or higher than 3.5 at 60 Mbps, whereas the results obtained in the immersive space at PSNC showed MOS scores higher than 3.5 at 85 Mbps, that is, a bitrate higher than in Deep Space 8K. An ANOVA analysis was conducted to investigate the influence of the environment on the perceived quality. The analysis concluded that, despite the higher MOS values produced in Deep Space, there was no statistically significant difference in the results at different bitrates between these two immersive spaces (see Section 3.2.2.4 for more details).

Table 10. 95% Confidence Interval - Deep Space 8K

| | 25 Mbps | 40 Mbps | 60 Mbps | 85 Mbps | 110 Mbps | 150 Mbps |
|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| Berlin | 4.16±0.29 | 4.39±0.28 | 4.35±0.28 | 4.61±0.23 | 4.45±0.25 | 4.55±0.27 |
| Follow | 4.00±0.37 | 4.23±0.26 | 4.45±0.23 | 4.68±0.23 | 4.65±0.20 | 4.52±0.25 |
| Island | 4.10±0.26 | 4.42±0.23 | 4.42±0.25 | 4.45±0.27 | 4.39±0.28 | 4.42±0.25 |
| MC2 | 3.97±0.35 | 4.42±0.25 | 4.61±0.23 | 4.71±0.19 | 4.58±0.23 | 4.55±0.23 |
| Cathedral | 2.55±0.50 | 2.84±0.47 | 3.45±0.45 | 3.48±0.42 | 3.81±0.38 | 4.10±0.36 |
| Singing | 2.97±0.41 | 3.74±0.34 | 3.74±0.42 | 3.81±0.36 | 4.00±0.28 | 4.16±0.27 |
| Average | 3.62±0.17 | 4.01±0.15 | 4.17±0.14 | 4.29±0.13 | 4.31±0.12 | 4.38±0.11 |

Figure 7 and Table 10 illustrate the obtained 95% CIs. Similarly to the 8K video wall, the error margins for Cathedral and Singing Sand are higher than those for the rest. We can also observe that the CIs corresponding to the Deep Space 8K environment get narrower as the bitrate increases in most the assessed video sequences at least until 85 Mbps.

The comments given by the participants are summarized below:

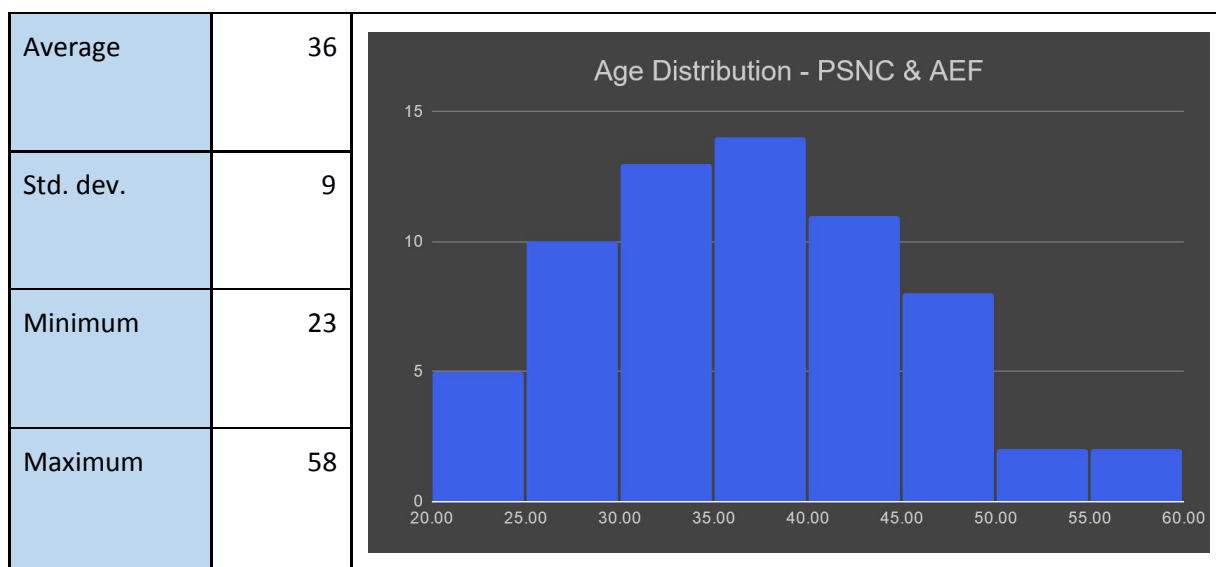
1. Several participants admitted that with time they started to see more. That they selected some spots to observe, where they could see the differences in quality (e.g. cables in Island in the Sky, tires of bicycles).
2. The most apparent differences could be seen in the Cathedral clip (especially on the floor of Cathedral as well as on the right section of Cathedral). Also in Singing Sand differences could be spotted in particles (getting blurry in some impaired videos).
3. For many viewers it was very difficult to spot any difference in the snow video. Some said the easiest way to spot difference was in observing moving images.
4. Many participants criticized the bad quality of original videos, having imperfections like jittering, stuttering, blurry images, etc., which made comparing the impaired to original video rather difficult.
5. Participants perceived the test as rather exhausting because of having to look closely at images. They showed much interest in getting an insight into the final test results.

The first three comments were somehow similar to those given by the PSNC's participants. In the fourth comment, according to the experimenter, the subjects referred especially to the "bad quality" of Island in the Sky II, Singing Sand, and Cathedral. With respect to the last comment, it can be understandable the fact that the subjects felt exhausted after a while, since a 1-hour session is perhaps too much for this type of experiment that requires a lot of concentration.

3.2.2.3 Global Results

In total, 65 people participated in the subjective test. Exactly 72.3% of the participants were men (47 persons of 65), whereas only 18 were women. By level of expertise, 75.3% (49 persons) were non experts and only 16 were experts. The combined age distribution is given in Table 11.

Table 11. Age Distribution - PSNC 8K video wall and Deep Space 8K



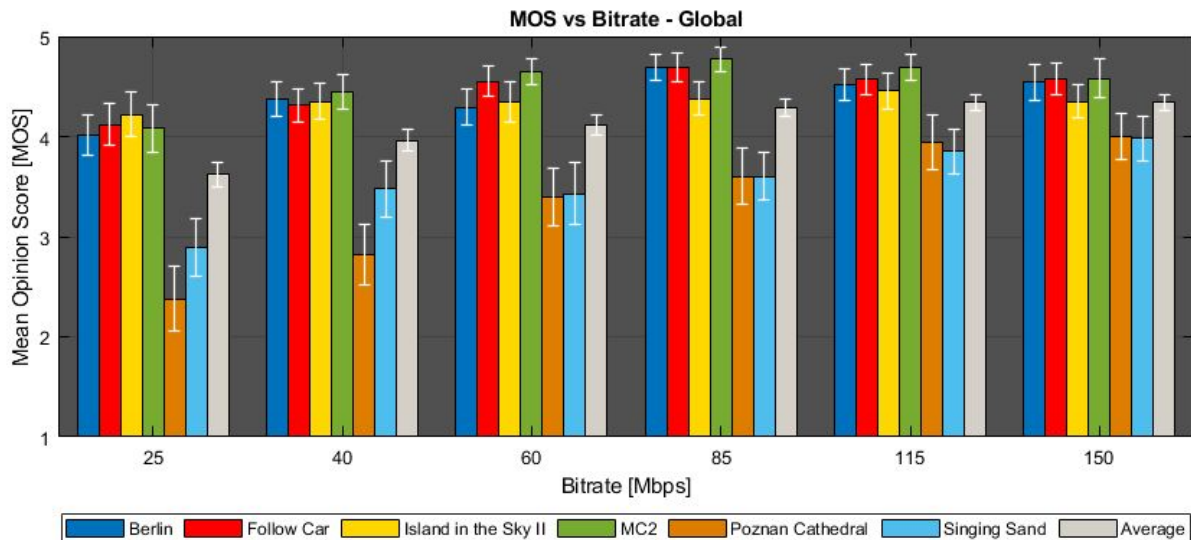


Figure 8. MOS versus Bitrate - PSNC 8K video wall and Deep Space 8K

Figure 8 shows the global results in terms of MOS at different bitrates for a total sample size of 65 participants. If we follow the recommendations from Sujito et al¹¹, the following two conditions should be satisfied to achieve broadcast quality HEVC video:

- The average MOS of the sequence under assessment should be 3.5 at least
- No sequence should have a MOS less than 3.0 (“slightly annoying”)

Based on these conditions, we can conclude that the minimum bitrate for Spin Enc Live to achieve **broadcast-grade quality 8Kp60 HEVC video is 60 Mbps** (minimum average MOS is almost 3.5), or 40 Mbps without considering special content, such as CGI or point-cloud video. In addition, according to the figure, the average MOS reaches the **saturation point at 85 Mbps**. If we consider special content only, we cannot ensure that the saturation point has been reached at 150 Mbps (at that bitrate the MOS is approximately 4.0). In order to find that point for **point-cloud content**, an additional informal subjective test was conducted by PSNC and SD and the results showed that bitrate points **higher or equal than 400 Mbps** produce high perceptual quality. More details about this test is reported in Section 4.3.6.

Table 12. 95% Confidence Interval - PSNC 8K video wall and Deep Space 8K

| | 25 Mbps | 40 Mbps | 60 Mbps | 85 Mbps | 110 Mbps | 150 Mbps |
|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| Berlin | 4.02±0.20 | 4.37±0.17 | 4.29±0.18 | 4.69±0.13 | 4.52±0.16 | 4.54±0.18 |
| Follow | 4.12±0.21 | 4.31±0.17 | 4.55±0.15 | 4.69±0.14 | 4.57±0.15 | 4.58±0.16 |
| Island | 4.22±0.22 | 4.35±0.18 | 4.34±0.20 | 4.38±0.17 | 4.46±0.18 | 4.35±0.17 |
| MC2 | 4.08±0.24 | 4.45±0.17 | 4.65±0.13 | 4.77±0.12 | 4.69±0.13 | 4.58±0.19 |
| Cathedral | 2.38±0.32 | 2.82±0.30 | 3.40±0.29 | 3.60±0.28 | 3.94±0.27 | 4.00±0.23 |

¹¹ Y. Sugito, S. Iwasaky, K. Chida, K. Iguchi, K. Kanda, X. Lei, H. Miyoshi, K. Kazui, *Video Bit-rate Requirements for 8K 120-Hz HEVC/H.265 Temporal Scalable Coding: Experimental Study based on 8K Subjective Evaluations*, APSIPA Transactions on Signal and Information Processing, 2015

| | | | | | | |
|---------|-----------|-----------|-----------|-----------|-----------|-----------|
| Singing | 2.89±0.29 | 3.48±0.28 | 3.43±0.31 | 3.60±0.24 | 3.85±0.22 | 3.98±0.22 |
| Average | 3.62±0.12 | 3.96±0.11 | 4.11±0.10 | 4.29±0.09 | 4.34±0.08 | 4.34±0.08 |

Figure 8 and Table 12 show the 95% CIs when adding the results produced in both immersive spaces. As we can see, the obtained error margins are lower than those shown in Tables 8 (PSNC video wall) and 10 (Deep Space 8K), because of the fact that sample size is approximately twice bigger. In many cases the error margin is lower than 0.20 except for Cathedral and Singing Sand in which it increases up to 0.30. In general, the CIs are at low bitrates higher than at high bitrates.

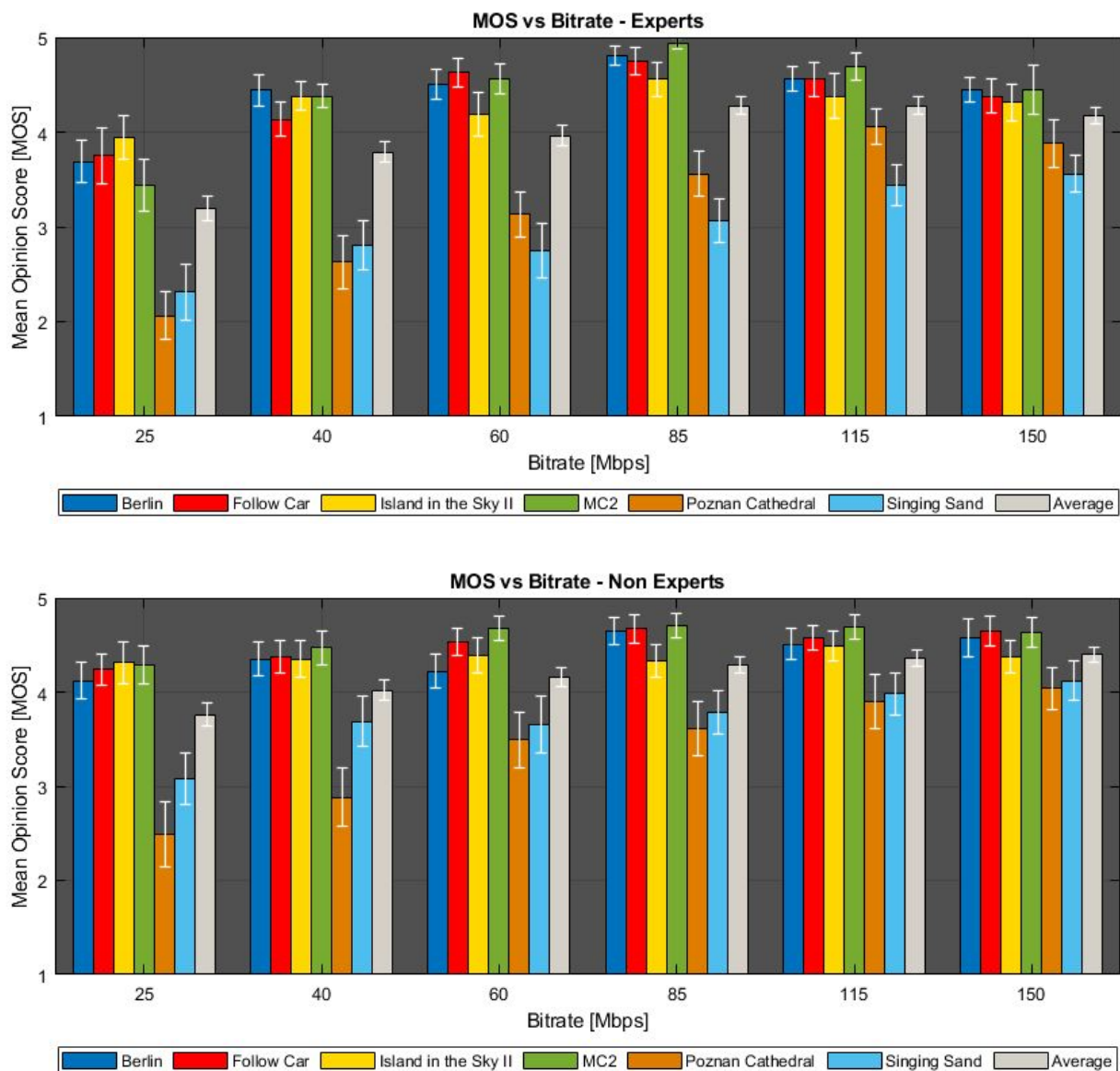


Figure 9. MOS-vs-bitrate comparison between experts (top) and non experts (bottom)

The results comparing people with different levels of expertise are illustrated in Figure 9. We can see that, in general, the **49 non-expert subjects gave higher scores than the 16 expert ones**, especially at low bitrates. In other words, the expert group was able to detect more quality degradations than the

non-expert group. This conclusion is in fact similar to that drawn by Sugito and Bertalmío¹², whose work proves that DSIS-based subjective tests with experts can better determine the lower threshold of image quality.

As a general conclusion, we think that the obtained scores were, in general, higher than initially expected for natural content, since in a pre-pilot test on a 8K monitor (see Section 3.2.1.4) noticeable differences between the original and the low-bitrate versions could be perceived by a video coding expert.

One reason can be that **the display technology (projector vs monitor/TV) might also influence on the perceived quality**, since projectors produce, in general, more blurred pictures and less contrast than LCD TVs or monitors. However, a subjective test using an 8K flat screen TV would be needed to confirm this hypothesis.

Another plausible reason can be that, even at the viewing distances recommended for the environments under evaluation, **the screens are perhaps too big for the users to really localize all the distortions**. From our point of view, the question about how one participant analyzed the videos on those big screens may have two possible answers:

1. **One behavior can be that the participant focused the central visual field only on the part of the picture in front of him/her.** In such a case, perhaps it was hard to spot coding artefacts in other parts of the picture, especially for those participants who sat at the extreme sides of the viewing line.
2. **Another behavior can be the fact that the subject tried to scan the whole picture for 10 seconds.** Although the videos were presented twice because of the DSIS-based evaluation method, maybe a 10-second duration is not enough for the participant to spot on the screen all the noticeable coding artefacts. Note that, according to Rec. P.913, stimuli ranging from 5 to 20 seconds in duration are allowed, but 8 to 10 seconds are recommended in order to prevent the temporal forgiveness effects. Moreover, although after some evaluations the participant got more and more familiar with the content and learned more about the coding errors, this effect is considered as a playback-ordering bias that could be minimized by means of the stimuli randomization effect described in Section 3.2.1.6.

3.2.3 Color Banding in Images with Gradient Ramps

3.2.3.1 Introduction

Display technologies have improved significantly in recent years. Not only the resolution has increased from 4K to 8K and frame rate from 60 Hz to 120 Hz, but also the bit-depth from 8-bit to 10-bit. Current state-of-the-art displays supporting a bit-depth of 10-bit are able to show pictures in very high quality with wide color gamuts and high dynamic range, as long as they have been generated in a 10-bit format at least. Nevertheless, in some particular cases this technology is still insufficient to display pictures at the maximum possible quality. One clear example is content presenting gradient color ramps. This type of content is indeed very sensitive to the well-known color

¹² Y. Sugito, M. Bertalmío, Non-experts or Experts? Statistical Analyses of MOS using DSIS Method, 2020 (accepted for publication)

banding effect¹³ when the bit-depth is too low. Color banding is typically present in skies, very dark and night scenes, and some CGI content.

Within the context of the Immersify project, an animation logo with smooth gradient ramps was created by AEF for showcase at different project-related events. This section describes all the work performed by AEF and SD towards the reduction of the color banding effect presented in this content (see Figure 10). Part of this work also included some informal subjective tests that were conducted to visually verify the enhancements created by AEF.

The section is divided into two parts. In the first part, we will describe the creation process of a first version of the content, as well as the encoding and playback tests carried out. In the second part, we will describe the work we carried out relative to an enhancement version of the animation logo.

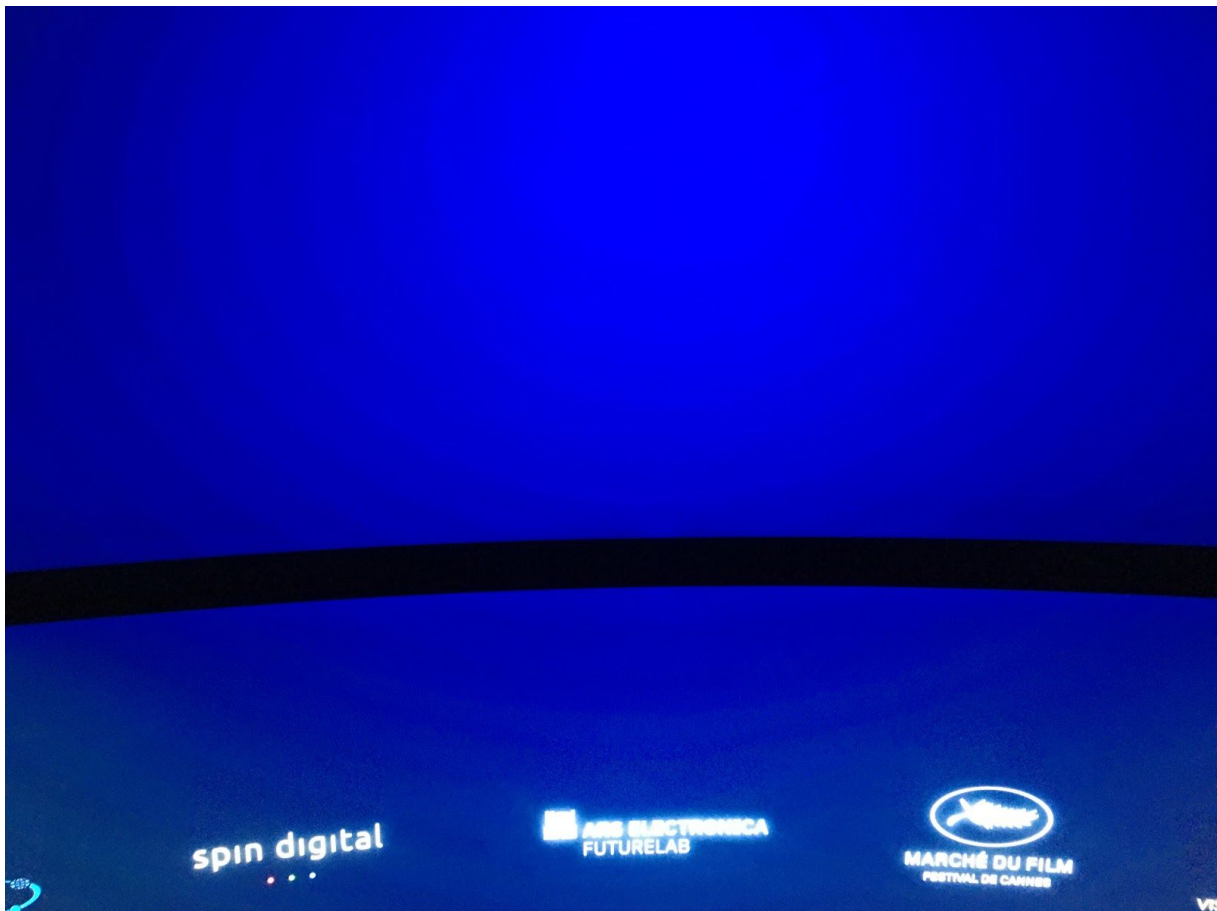


Figure 10. Immersify logo animation presenting color banding in the background

3.2.3.2 *Immersify Logo Animation: Version 1*

First approach: 10-bit export format

The Immersify logo animation was produced in Adobe After Effects CC as a mixture between procedural 3D particle systems and the means of traditional 2D key-frame animation. It was rendered as a 8K 60 fps non-stereoscopic version for general distribution as well as a 8K 30 fps/eye

¹³ [W. Sawalich, Identifying and Repairing Banding, Dec. 12, 2016.](#)

stereoscopic version for 3D use-cases (e.g. Ars Electronica Deep Space or HMDs). With a duration of 20 seconds it consists of 1200 individual frames.

The issue of color banding was encountered in dark color gradients after encoding the first full render from PNG sequence to video file. Banding occurs in gradients composed of too little information as visible steps in tonal value between the starting and the end colour. Although the initial production project is set to 32-bit per color channel inside Adobe After Effects CC with just 8-bit per channel the rendered PNG sequence had too little information to begin with.

In a first attempt to resolve the issue CINEON was chosen as an image sequence export format. It can store 4:4:4 chroma format and up to 10-bit of color information per channel. The rendered CINEON image sequence was then handed over to SD for encoding.

Informal subject test

To verify the quality of the Immersify logo animation, an informal subjective test was performed in the SD lab using a Samsung 8K TV and the media player integrated in the TV. Spin Player could not be used in this experiment, because the HDMI input contentor does not support 8Kp60 signals.

Spin Enc Offline v1.8 was the software used to pre-process and encode the CINEON image sequence. In the pre-processing stage the images were down converted from 4:4:4 to 4:2:2 (contribution format) as well as from 4:4:4 to 4:2:0 (distribution format). Afterwards, the broadcast encoding configuration (CBR rate control, GoP size of 16 frames, and intra period of 64 frames) was used to encode these two versions at two different bitrates: 120 Mbps and 200 Mbps.

The conclusion drawn from the experiment was that, independent of the bitrate and chroma format, color banding could be noticed on the TV but it was not annoying. However, when this animation was projected on a cinema screen for a demo at Cannes XR 2019 using the Digital Projection 8K projector, the artefact turned out to be very visible, even annoying for some visitors who attended the demo. It seems that there is a correlation between the amount of brightness and contrast a display technology is able to emit and the strength of color banding.

Second approach: dithering and 12-bit export format

In a new approach dithering was used within the gradient-ramps in Adobe After Effects CC to introduce scattered blending of the respective colors and their tonal values. Dithering helps to maintain the appearance of smooth transitions after encoding. The image sequence export was changed to the Digital Picture Exchange standard format DPX to allow rendering in 10-bit, 12-bit or up to 16-bit color channels. The newly rendered 12-bit DPX image sequence was then handed over to SD for encoding.

Informal subjective test

This new master file was down converted to 4:2:0 10-bit and then encoded at 120 Mbps with the abovementioned configuration. Another informal subjective test was performed in the SD lab on the 8K TV. The test concluded that, even with this dithering-based approach, color banding still appeared, basically because of the sub-conversion from 12-bit to 10-bit in the encoder. That is, 10 bit is not enough bit-depth to represent this kind of images with gradient ramps.

In order to verify if 12-bit encoding can have some benefit, the content was encoded again but without any bit-depth conversion. As the media player integrated in the 8K TV does not support 12-bit HEVC decoding, the content was tested on a layout of 2x2 4K 10-bit monitors using a PC with Spin Player. In this experiment we could notice color banding mainly because of the fact that the GPU converted to 10-bit for final display on the 4K monitors. It is also worth mentioning that the banding effect was even more visible when using dark colors (dark blue, black...).

3.2.3.3 *Immersify Logo Animation: Version 2*

Final approach

As banding still occurred in the dark color gradients another version of the Immersify logo animation was created. In the Adobe After Effects CC project all gradient ramp color values were brightened up to reduce the dark areas overall. This process introduced better color information distribution within the areas prone to banding artefacts. Finally a 1% noise layer was added on top to further help breaking up the gradient color values.

The Immersify logo animation was then rendered as 8K 60 fps non-stereoscopic DPX 4:4:4 10-bit image sequence and handed over to SD for encoding. A second 8K 30 fps/eye stereoscopic DPX image sequence was encoded at Ars Electronica Futurelab and tested in Deep Space.

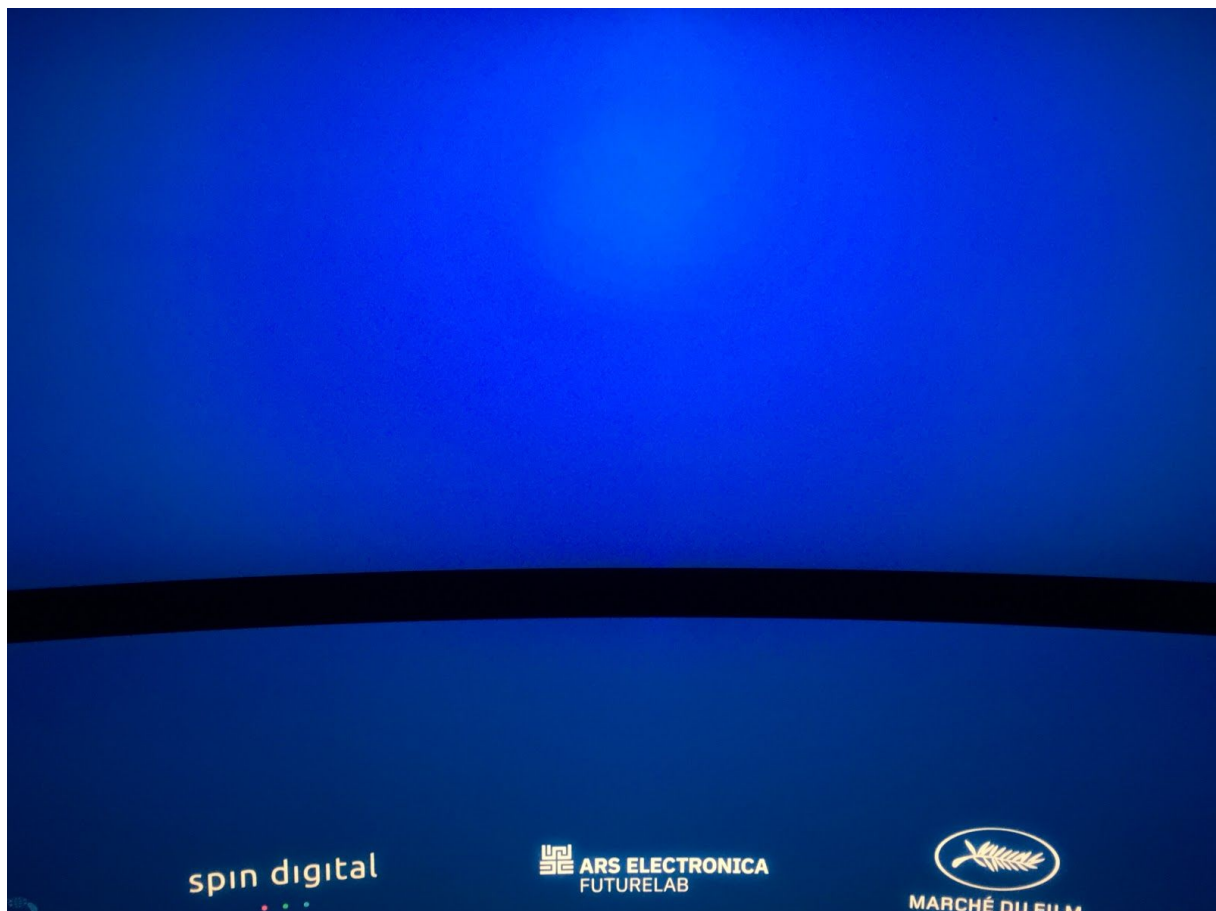


Figure 11. Immersify logo animation with reduced color banding

The new version was downscaled to 4:2:0 and encoded at 120 Mbps using the broadcast encoding configuration. The visual quality of this content was better than of the previous approach (see Figure 11), although some banding artefacts were still visible in the transition from dark blue to black as well as in the fade-out at the end.

With this final approach we were able to reduce color banding to a satisfactory extent. In order to remove this effect entirely 12-bit displays would be needed, unfortunately such a technology is not available yet.

3.2.4 Encoder Settings for Point-cloud Video

3.2.4.1 Introduction

In Immersify we are dealing with content in which color perceptibility potentially deviates from standard camera content, especially in cases where the textures are highly discontinuous (e.g. the particles in Singing Sand or the dot structure in point cloud renderings from laser scanings). At the same time, this type of content is more challenging for traditional video encoders, such as HEVC or H.264, due to its non-continuous structure. Indeed, coding tools work efficiently in natural content, where *spatial and temporal redundancy* is very high. Spatial redundancy refers to pixels that are spatially close to each other and have similar values. Temporal redundancy means that consecutive pictures are almost identical. In the case of computer-generated content made of dots or particles separated by holes this spatio-temporal redundancy drops significantly and, as a result, a significant amount of bits is required to keep a certain level of quality of encoded picture. However, the case of Singing Sand is not so critical, because this content presents big portions of background with a uniform color and, hence, the bit budget per picture is mainly dedicated to the encoding of the particles.

PSNC has developed a complete production workflow that converts the point cloud from laser scanning to a sequence of digital pictures. In our experiments we used FARO Focus 3D X330 laser scanner for all of the scans. We mostly used FARO SCENE software for stitching and editing scans for removing and cleaning unnecessary points, as it integrated seamlessly with the FARO scanner. For final post-production and rendering to video in 8K, 16K, 360° and 3D format, we used *CloudCompare* for rendering Cathedral and *Blender* for further projects including Fallout Shelter.

In the particular case of *Blender*, this software allows the user to select the point size in pixels. We believe that this size might also influence the compression efficiency. Intuitively, larger point sizes should produce higher compression efficiency, because a reduction in spatial frequency can lead to a reduction in the number of transform coefficients needed to represent the content with acceptable quality.

The objective of this task is twofold: 1) to find the most favorable point size from the HEVC encoding point of view; and 2) to find the recommended encoder configuration to generate high-quality point-cloud HEVC video).

3.2.4.2 Point Size Selection

In order to find the most appropriate point size, an informal subjective test was performed in the SD lab with different point sizes from 2 pixels to 30 pixels. The point-cloud content called *Fallout Shelter*

(more information can be found in report D5.2 “Report on Immersive Content Production”, was used for this purpose. This content was encoded with Spin Enc Offline v1.9-dev at two bitrates, 200 and 400 Mbps, in CBR mode, and then screened on an 8K monitor. In particular, this encoder version introduces a new rate control algorithm that enables higher quality consistency than the previous versions. It relies on an advanced lookahead algorithm with a longer analysis window that is able to predict much better the amount of bits that the next 1-second video will yield.

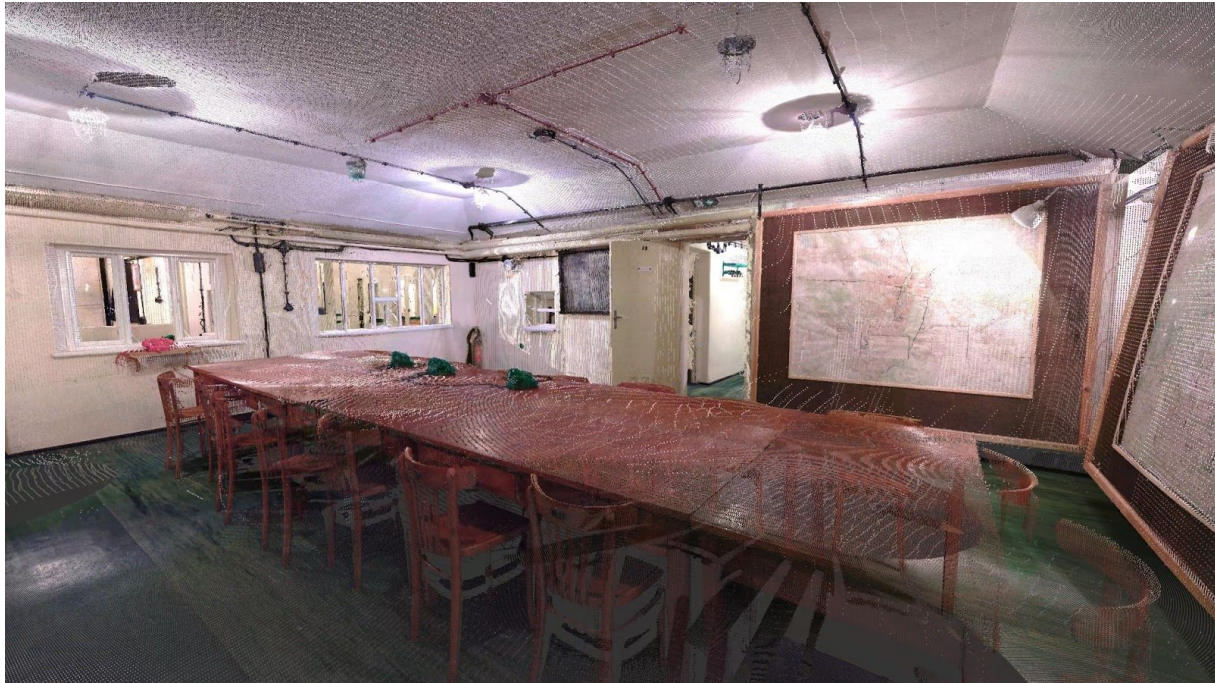


Figure 12. Fallout Shelter rendered to video in 8K resolution with a point size of 7 pixels

The test concluded that the optimal point size for 8K video in terms of compression efficiency ranged from 5 pixels to 10 pixels. A point size of 2 pixels led to significant coding artefacts. On the other hand, point sizes bigger than 10 pixels produced better compression efficiency, but the resulting pictures looked like an “impressionist painting” and very artificial from the visual point of view. Based on this feedback, a size of 7 pixels was finally chosen by PSNC for 8K (see Figure 12). Another subjective experiment was also performed with a 360° CMP version rendered in 12960x8640 resolution and point size of 7 pixels. The experiment concluded that this point size also produced good quality of the encoded video.

3.2.4.3 Encoder Configuration

CBR rate control is recommended to keep the bitrate very stable over time when the bitrate range is close to the maximum capacity of the playback machine. For point-cloud content, GoP sizes from 4 to 16 and Intra periods from 64 (1-second) to 128 (2 seconds) produce good visual quality when using Spin Enc v1.9-dev with its new rate control approach.

An informal test was performed by PSNC to find a good bitrate point that guarantees high-quality point-cloud 8K video. To this end several versions of Fallout Shelter with a point size of 7 pixels were created at different bitrates from 200 Mbps to 1000 Mbps and then displayed on the 8K wall (WxH: 6m x 2.8m) at PSNC using Spin Player. According to the PSNC team, those versions encoded at

bitrates equal or higher than 400 Mbps were subjectively similar to each other. The visual quality of Cathedral encoded at bitrates higher than the ones used for the subjective quality assessment was also verified. Although this content was created with a different post-production workflow (CloudCompare), this informal test also concluded that 400 Mbps is a good value for this content.

3.2.5 Encoder Settings for “Immersive Minimalism”

3.2.5.1 Introduction

Within the programme Starts Residences, PSNC hosted Thersa Schubert’s residency under the title “Immersive Minimalism” for the creation of a 10-minute film in 4K and 8K resolution¹⁴. In this project a Cellular Automata (CA), a set of rules typically used for simulating patterns in nature and physical behaviors, has been used to create patterns based on cells (or pixels) that are self-organized over time after several iterations of the CA. This project was presented at Ars Electronica Festival 2019.

In the next section we will describe the collaborative work performed by Theresa and SD towards the creation and encoding of the master file in HEVC at the maximum possible quality.

3.2.5.2 Encoder Configuration

From the purely encoding point of view, Immersive Minimalism behaves like a white noise. Due to its “noisy” nature, this content is also very challenging for traditional encoders, since the encoding tools have not been designed to encode noise in an efficient way. As a result, this content also requires an especial encoder configuration to achieve a certain level of quality of encoded video.

The encoding and subjective tests conducted with the aim of finding the best encoder configuration for this content were divided into two phases: *learning phase and final encoding phase*.

In the first phase some preliminary subjective tests were performed in the SD lab using a Panasonic 4K TV (only the top-left quarter was displayed) and in the PSNC lab using a Sharp 8K TV. A few excerpts of Immersive Minimalism were carefully selected and exported in DPX 4:4:4 10-bit format at 30 Hz. The master files were consequently encoded using *Spin Enc Offline v1.8* with different configurations. The following conclusions were drawn:

- The 4:2:2 and 4:4:4 formats are not needed for this content, since it presents a very few range of colors.
- A bit-depth of 10-bit can be used, since compression-wise it is more efficient than 8-bit.
- The bitrate for 8K should be at least 1000 Mbps.
- The encoder configuration should be CBR and IPPP (GoP = 1). Higher GoP sizes produced a visible flickering effect due to the fact that hierarchical patterns encode the pictures inside the GoP with different qualities.
- The new rate control implemented in Spin Enc Offline 1.9-dev might help to produce a smoother evolution of the video quality.

¹⁴ [T. Schubert, Immersive Minimalism at Ars Electronica Festival, START Residences Blog, 2019.](#)

- A frame rate of 30 Hz resulted in an annoying flickering (or stroboscopic) effect. This effect was also verified by PSNC on an 8K TV.

In the final encoding phase the complete version of Immersive Minimalism was exported in DPX 4:4:4 10-bit at 60 Hz and then encoded with Spin Enc Offline v1.9-dev (at that time this version was under development) at 1000 and 1500 Mbps. Informal subjective tests were conducted on two displays: a Panasonic 4K TV at SD (only a 4K portion was displayed), a Sharp 8K TV at PSNC. The playback machines used for these tests were powerful enough to process 8Kp60 video at up to 1500 Mbps. The experimental results were concluding: the quality produced by the version at 1500 Mbps was quality similar to that at 1000 Mbps. We could also check out that the new encoder with a better rate control produced better quality consistency than the old one included in Spin Enc Offline v1.8.



Figure 13. Immersive Minimalism on Deep Space 8K display (floor and wall projections)

A final informal test was also conducted in the immersive environment at Ars Electronica. Since the resolution in pixels of the Deep Space wall is, due to edge blending, 6467x3830, the video was downscaled to that screen resolution and then encoded at 750 Mbps rather than at 1000 Mbps, as the ratio between the 8K resolution and the Deep Space resolution is almost equal to 0.75. According to the AEF team, the 6467x3830 content at 750 Mbps could be seen without any noticeable coding artefact. For the final screening at Ars Electronica Festival, the floor projection was generated as well (see Figure 13).

3.2.6 Advances over the State of the Art

Although some papers on subjective quality assessment of video codecs have been published, such as Katsenou et al.¹⁵ and Dias et al.¹⁶, the only one that specifically focuses on 8K is Sugito et al.¹⁷. The aim of that paper work is similar to that of Task 5.3, which is to find the best bitrate configuration for the 8K broadcast use case. The authors of the paper used a hardware encoder designed for broadcast applications¹⁸. In our formal experiments we assessed the subjective quality produced by a state-of-the-art 8K real-time software encoder which is also tailored for broadcast streaming, as well as for live internet.

Moreover, unlike the previously mentioned studies, the set of test sequences also included very special content such as point-cloud video. As far as we know, no other study on subjective quality assessment has considered this type of content. Therefore, owing to all these particular testing conditions, the outcome of Task 5.3 may also be of special interest for the scientific community and media industry.

3.3 Status

3.3.1 Expected Results

The expected results for Task 5.3 is:

“Guidelines for encoding and decoding immersive content, as well as media production workflows experiences. As a best practice reference, the information will also be published on the project website.”

In addition, in order to come up with recommended encoding parameters a quality testing phase of the real-time software encoder had to be conducted as promised in the GA:

“Specific quality-rate points will be selected according to the demands of end-users and the system requirements for the envisioned demonstrations (Tasks 5.4 and 5.5). Optimal configurations of the encoder will be found for maximizing the subjective quality given a target bitrate, while at the same time guaranteeing real-time operation of the decoder. An informal subjective test will be prepared on the target display environments in order to validate the improvements implemented in the encoder in terms of subjective quality.”

¹⁵ A. V. Katsenou, F. Zhang, M. Afonso and D. R. Bull, *A Subjective Comparison of AV1 and HEVC for Adaptive Video Streaming*, 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 2019, pp. 4145-4149.

¹⁶ A.S. Dias, S. Blasi, F. Rivera, E. Izquierdo, M. Mrak, *An Overview of Recent Video Coding Developments in MPEG and AOMEDIA*. INC Conference 2018

¹⁷ Y. Sujito, S. Iwasaky, K. Chida, K. Iguchi, K. Kanda, X. Lei, H. Miyoshi, K. Kazu, *Video Bit-rate Requirements for 8K 120-Hz HEVC/H.265 Temporal Scalable Coding: Experimental Study based on 8K Subjective Evaluations*, APSIPA Transactions on Signal and Information Processing, 2019

¹⁸ Y. Sugito, S. Iwasaki, K. Chida, K. Iguchi, K. Kanda, X. Lei, H. Miyoshi, Y. Uehara, *8K 120-Hz real-time video codec*, in *2019 NAB Broadcast Engineering and Information Technology Conference (BEITC)*, Las Vegas, NV, USA, 453–457, 2019

3.3.2 Obtained Results

In the GA informal subjective tests were envisioned for finding the best encoder configuration, especially the bitrate value. However, we finally decided to perform a formal subjective test on two target display environments using content created or used within the context of Immersify. So the design, implementation, and execution of formal tests can be considered as an exceeding result.

Finally, more subject tests were also conducted in this task with the objective of finding the best encoding configuration of special content used in the project, such as the Immersify logo animation content, point-cloud video (Cathedral and Fallout Shelter), and Immersive Minimalism.

4 Encoding and Playback Settings

In this section we provide the encoder and playback settings that we recommend based on the results given in the previously described task on quality assessment. These recommended settings will be divided into the use cases envisioned in the Immersify project.

We present five use cases that are typical and characteristic for state-of-the-art high-resolution immersive content. Our use cases include the playback of different types of immersive content such as 8K 2D, 8K 3D, high resolution 360°, point-clouds, and interactive video playback

For each use case we also provide the exact technical specifications of the video and the technical environment in which the test has been performed. These specifications, specially for video playback, define the minimum technical requirements needed for supporting the target use case.

It is also worth highlighting that all the recommendations that we will provide in this section ensure reliable encoding, transmission, and playback if the tools developed in the Immersify project are used. These tools include: HEVC encoder (real-time and offline) and media players based on Spin Digital SDK (Spin SDK).

4.1 8K 2D Video

4.1.1 Description

The playback of an 8K 2D video content on a flat surface is the most common and the simplest case of immersive video content, whereas 8K 2D video content will be played with projectors or TVs and monitors. In our study we use 8K video content with simple interaction options (play, pause, seek, zoom).

4.1.2 Technical Requirements

This use case is defined for 8K 2D videos at 60 fps. All the videos are converted before encoding to chroma subsampling 4:2:0 and 10-bit. Video encoding is based on HEVC Main 10 profile. The audio is encoded either as a stereo stream or as a 5.1 channel stream. The table below gives more detailed information on technical specifications of this use case.

| | | |
|-------|-----------------------------|------------------------------------|
| Video | Target display | Deep Space 8K, PSNC 8K wall, 8K TV |
| | Resolution and frame rate | 8Kp60 |
| | Chroma format and bit depth | 4:2:0, 10-bit |
| | Input projection | Plane |
| | Output projection | Plane |
| | Content source | Pre-encoded file |
| | Output rendering device | GPUs |

| | | |
|-------------|----------------------------|-------------------------|
| | Single / Multiple machines | Single |
| Audio | Input | 2.0, 5.1 |
| | Output | 2.0, 5.1 |
| Interaction | Operations | Play, pause, seek, zoom |

4.1.3 Encoder Settings

| | | |
|-----------|---------------|--|
| Video | Encoder | Spin Enc Offline (spinfompeg) |
| | Coding scheme | HEVC Main 10 Profile (4:2:0 10-bit) |
| | GoP size | 16 frames (hierarchical) |
| | Intra period | 64 frames |
| | Rate control | CBR |
| | Video bitrate | 60 - 120 Mbps |
| Audio | Encoder | Native FFmpeg AAC encoder (spinfompeg) |
| | Rate control | VBR (default) |
| | Bitrate | 2.0: 160 Kbps 5.1: 480 Kbps |
| Container | | MP4, MPEG2-TS |

4.1.4 Playback Settings

| | |
|--------------|--|
| Media player | - Spin Player - Custom player based on Spin HEVC Decoder (Spin SDK) |
| OS | Windows 10 64-bit |
| CPU | Intel Core i9-9900x (10 cores), or AMD Ryzen 9 3900X (12 cores) |
| GPU | NVIDIA Quadro P4000, or AMD Radeon Pro WX 7100 |
| Memory | 16 GB (4x 4 GB, DDR4 3200) |

4.2 8K 3D (Stereoscopic) Video

4.2.1 Description

The playback of 8K 3D (stereoscopic) video content on a flat surface is an extension of the previous case. In Immersify we support stereoscopic content using active 3D projection (e.g. Barco F50 WQXGA¹⁹ for the PSNC 8K wall). Active projection relies on active glasses that alternate between opened and closed states, so that each eye sees a different picture. The video rendering system should provide the projector with 2 images, one for each eye, either by increasing the frame rate with a frame interleaving method (going from 60 to 120 fps) or by increasing the spatial frame resolution using frame packed formats (and from 8Kx4K to 8Kx8K).

In our applications we use 8K 3D video content with simple interaction options (play, pause, seek, zoom).

4.2.2 Technical Requirements

We support 8K 3D playback by using frame packing: side-by-side or top-bottom, where for top-bottom layout the top part of the picture corresponds to the left view and the bottom part to the right view. The total resolution in the case of top-bottom is 8Kx8K (7680x7680 px).

The videos are converted before encoding to a frame packed layout, and also converted to chroma subsampling 4:2:0 and 10-bit. The frame-packed videos are then encoded using an HEVC encoder without special support for 3D (i.e. HEVC MVC or 3D extensions).

The audio is encoded either as a stereo stream or as a 5.1 channel stream. The table below gives more detailed information on technical specifications of this use case.

| | | |
|-------|-----------------------------|---|
| Video | Target displays | Active 3D displays: <ul style="list-style-type: none"> - PSNC 8K wall - Deep Space 8K |
| | Resolution and frame rate | 8Kx8Kp60 |
| | Chroma format and bit depth | 4:2:0 10-bit |
| | Input projection | Plane |
| | Output projection | Plane |
| | Stereopair layout | Top-bottom Side-by-side |
| | Content source | Pre-encoded file |
| | Output rendering device | GPUs |
| | Single / Multiple machines | Single |

¹⁹ <https://www.barco.com/en/product/f50-wqxga>

| | | |
|-------------|------------|-------------------------|
| Audio | Input | 2.0, 5.1 |
| | Output | 2.0, 5.1 |
| Interaction | Operations | Play, pause, seek, zoom |

4.2.3 Encoder Settings

| | | |
|-----------|---------------|--|
| Video | Encoder | Spin Enc Offline (spinffmpeg) |
| | Coding scheme | HEVC Main 10 Profile (4:2:0 10-bit) |
| | GoP size | 16 frames (hierarchical) |
| | Intra period | 64 frames |
| | Rate control | CBR |
| | Video bitrate | 120-240 Mbps |
| Audio | Encoder | Native FFmpeg AAC encoder (spinffmpeg) |
| | Rate control | VBR (default) |
| | Bitrate | 2.0: 160 Kbps 5.1: 480 Kbps |
| Container | | MP4, TS |

4.2.4 Playback Settings: Minimum Requirements

| | 8K 3D |
|--------------|-------------------------------------|
| Media player | Spin Player |
| OS | Windows 10 64-bit |
| CPU | Intel Xeon Platinum 8168 (24 cores) |
| GPU | 2x NVIDIA Quadro P4000 |
| Memory | 32 GB (4x 8 GB, DDR4 3200) |

4.2.5 Playback Settings: Immersive Environments

| | PSNC 8K Video Wall | Deep Space 8K |
|--------------|--------------------|---|
| Media player | Spin Player | <ul style="list-style-type: none"> - Unity3D player (using native C++ Immersify plugin) - OpenGL Core Player (using Spin SDK) |

| | | |
|--------|---|---|
| OS | Windows 10 64-bit | Windows 7 SP1 64 Bit |
| CPU | 2x Intel Xeon Platinum 8168 (2x 24 cores) | 2x Intel Xeon CPU ES-2687W (2x16 cores) |
| GPU | 2x NVIDIA Quadro P5000 | 4x Nvidia Quadro P6000 |
| Memory | 192 GB | 64 GB (8x 8 GB, DDR4 2400) |

4.3 High-resolution 360° Playback

4.3.1 Description

This use case shows the playback of 360° video content from omnidirectional cameras on a flat or curved screen. 360° videos use a special type of projection on a spherical field of view to map it onto a flat surface. There are different types of projection and the ones supported by Spin Player are: Equirectangular (ERP) and CubeMap (CMP). Viewers of a 360° video experience an immersive space that is mapped from an area of the source image. It means, the source image for a frame is not shown entirely, but only as a part (based on the viewer's perspective). This view can change through interaction for example via a game-controller or mouse and/or keyboard.

4.3.2 Technical Requirements

The resolution of the test videos is 8K or higher (depending on the original resolution of the input video) and the frame rate ranges from 24 to 60 fps. This type of content from cameras is generally very static and, therefore, it does not require high bitrate for high-quality compression. The videos shall be pre-encoded with 4:2:0 chroma subsampling and 10-bit.

The media for this use case contains 3rd order ambisonics audio (ACN/SN3D) and the output can be either standard multi-channel format (5.1, 7.1, 22.2, 24.1) or custom layout (e.g. PSNC 24.1). If the audio piece includes extra Low-frequency Effects (LFE) files for the subwoofer channels, these tracks shall be aggregated to the ACN/SN3D tracks in a new file. For example, if the ambisonics file is 3rd order and there is only one LFE signal, the first 16 tracks will be ambisonics and the last track will be LFE.

| | | |
|-------|-----------------------------|---|
| Video | Target display | PSNC 8K wall, 8K TV, video walls, Deep Space 8K |
| | Resolution and frame rate | 8Kp30/60, 12p30/60, 16p30/60 |
| | Chroma format and bit depth | 4:2:0, 10-bit |
| | Input projection | ERP, CMP |
| | Output projection | Plane: rectilinear or curved |
| | Content source | Pre-encoded file |
| | Output rendering device | GPUs |

| | | |
|-------------|----------------------------|---|
| | Single / Multiple machines | Single |
| Audio | Input | 3rd order ambisonics (ACN/SN3D format) |
| | Output | Multi-channel layouts: 5.1, 7.1, 22.2, custom |
| Interaction | Operations | Play, pause, seek, zoom, rotation, re-centering |
| | Device | XBOX controller, mouse |

4.3.3 Encoder Settings

| | | |
|-----------|---------------|--|
| Video | Encoder | Spin Enc Offline (spinffmpeg) |
| | Coding scheme | HEVC Main 10 Profile (4:2:0 10-bit) |
| | GoP size | 16 frames (hierarchical) |
| | Intra period | 64 frames |
| | Rate control | CBR |
| | Bitrate | - 8Kp30/60: 50-75 Mbps - 12Kp30/60: 100 - 150 Mbps - 16Kp30/60: 200 - 250 Mbps |
| Audio | Encoder | - Native FFmpeg AAC (spinffmpeg), or - Libopus (spinffmpeg) |
| | Rate control | VBR (default) |
| | Bitrate | 2048 Kbps |
| Container | | - HEVC & AAC: MP4 - HEVC & Opus: MKV |

4.3.4 Playback Settings

| | 8Kp60 | 12Kp60 | 16Kp60 |
|--------------|---|-------------------------------------|---|
| Media player | Spin Player | | |
| OS | Windows 10 64-bit | | |
| CPU | Intel Core i9-9900x (10 cores) | Intel Core i9-10980XE (18 cores) | 2x Intel Xeon Platinum 8260 (2x24 cores) |
| GPU | Up to 4x NVIDIA Quadro P4000 (the number of GPUs is determined by the number of output displays) | | |

| | | | |
|--------|-----------------------------|-----------------------------|--------------------------------|
| Memory | 16 GB 4x 4 GB, DDR4 3200 | 32 GB 4x 8 GB, DDR4 3200 | 96 GB (12x 8 GB, DDR4 2933) |
|--------|-----------------------------|-----------------------------|--------------------------------|

4.4 Point-cloud Video Playback

4.4.1 Description

Point-cloud video contents are rendered from point data. This data is commonly generated by Laser 3D scanners. The data consists of a set of points in space representing the surfaces in the 3D space.

Each point contains information about its location and the RGB color in the three-dimensional space. Compression of a video consisting of point-cloud data can be seen as challenging for the state of the art compression algorithms like HEVC. This is due to the empty spaces that occur between the pixels representing single points in the 3D space so that the dependencies between frames (e.g. for motion detection) decrease. Due to this, point-cloud content has to be encoded at higher bitrates than other types of video content, and the playback machine should be powerful enough to be able to handle such data rates.

4.4.2 Technical Requirements

The original point-cloud data is rendered with the Cloud Compare or Blender software to a CubeMap projection format as an image sequence, where the test videos are encoded from. Similar to the aforementioned high-resolution 360° video content, the audio was encoded in 3rd order ambisonic. For the CMP video file any resolution up to 13K (12960x8640 pixels) and frame rate up to 60 Hz can be used for achieving high-quality HEVC video. However, 50 or 60 fps is highly recommended when the point cloud render also includes trajectories with camera motions, otherwise video playback will become very choppy.

The bitrates that we recommend for 8K point-cloud content guarantee high-quality video without requiring maximum CPU utilization on a 48-core playback machine, which is the system that we have been using for demonstrations during the project time. Higher bitrates can be achieved when using more powerful systems, based on next generation CPUs with up to 64 cores per socket.

The bitrates for 12960x86840 video are mainly limited by the maximum processing capacity of the playback machine. Although higher bitrates would be desirable for achieving high-quality video playback, the ones specified in the table on encoder settings produce good quality video with the recommended playback workstation.

| | | |
|-------|-----------------------------|---|
| Video | Target displays | PSNC 8K wall, 8K TV, video walls, Deep Space 8K |
| | Resolution and frame rate | 8Kp30/60, 12Kp30, 16Kp30 |
| | Chroma format and bit depth | 4:2:0, 10-bit |
| | Input projection | Plane, CMP |
| | Output projection | Plane: rectilinear or curved |

| | | |
|-------------|----------------------------|---|
| | Content source | Pre-encoded file |
| | Output rendering device | GPUs |
| | Single / Multiple machines | Single |
| Audio | Input | 3rd order ambisonics (ACN/SN3D format) |
| | Output | Multi-channel layouts: 5.1, 7.1, 22.2, custom |
| Interaction | Operations | Play, pause, seek, zoom, rotation, re-centering |
| | Devices | XBOX controller, mouse |

4.4.3 Encoder Settings

| | | |
|-----------|---------------|---|
| Video | Encoder | Spin Enc Offline (spinffmpeg) |
| | Coding scheme | HEVC Main 10 Profile (4:2:0 10-bit) |
| | GoP size | 4 frames (hierarchical) |
| | Intra period | 64 frames |
| | Rate control | CBR |
| | Bitrate | <ul style="list-style-type: none"> - 8Kp25/30: 200 Mbps - 8Kp50/60: 400 Mbps - 13Kp25/30: 800 Mbps - 13Kp50: 650 Mbps (limited by the playback machine) - 13Kp60: 500 Mbps (limited by the playback machine) |
| Audio | Encoder | <ul style="list-style-type: none"> - Native FFmpeg AAC (spinffmpeg), or - Libopus (spinffmpeg) |
| | Rate control | VBR (default) |
| | Bitrate | 2048 Kbps |
| Container | | <ul style="list-style-type: none"> - HEVC & AAC: MP4 - HEVC & Opus: MKV |

4.4.4 Playback Settings

| | PSNC 8K wall, 8K TV, video walls | Deep Space 8K |
|--------------|----------------------------------|---|
| Media player | Spin Player | <ul style="list-style-type: none"> - Unity3D player (using native C++ Immersify plugin) - OpenGL Core Player (using Spin SDK) |
| OS | Windows 10 64-bit | Windows 7 SP1 64 Bit |

| | | |
|--------|---|--|
| CPU | 2x Intel Xeon Platinum 8260 (2x 24 cores) | 2x Intel Xeon CPU ES-2687W (2x 16 cores) v3 @ 3.30 Ghz |
| GPU | Up to 4x NVIDIA Quadro P4000 (depending of the number of output displays) | 4x Nvidia Quadro P6000 |
| Memory | 96 GB (12x 8 GB, DDR4 2933) | 64 GB (8x 8 GB, DDR4 2400) |

4.5 Interactive Video Playback

4.5.1 Description

The idea behind interactive video playback is to mix and get the best out of two worlds: The good performance from videos and the flexibility of real time rendering. A viewer is supposed not to recognize that a video is seen, but to think that the content is completely rendered in real-time.

To demonstrate this idea, we created an application for the Deep Space 8K called “the Great Pyramid”. In this application, a user sees a tour around and through the pyramid of Gizeh, visualized via point cloud scans and presented as interactive 360° video. While the 360° video is presented, a guide can control the view direction of the camera and is able to pause and resume the video at any point. The audio track always keeps in sync with the video content. At some points in the story, the user can decide where to move next, e.g. in the center of the pyramid, the user can choose (via buttons, selectable via floor tracking) to go up to the King's Chamber or down to the Subterranean Chamber.

4.5.2 Technical Requirements

Technically seen, we use several 360° videos that can be stitched together at their beginning & end, so that a user cannot visually see that the video was switched. The playback in Unity allows us to rotate the camera view and blend in additional content (e.g. a title or a mini-map) in real-time.

The usage of point cloud videos is of special interest in this case. The 360° video presents billions of points that cannot be rendered in real-time with currently available hardware. Nevertheless, the user gets the feeling of seeing a real-time application.

| | | |
|-------|-----------------------------|---|
| Video | Target displays | Deep Space 8K, various HMDs |
| | Resolution and frame rate | 10K x 10K, 12K x 12K, 30 fps, progressive |
| | Chroma format and bit depth | 4:2:0 8-bit |
| | Input projection | CubeMap |
| | Output projection | Plane (Wall and Floor) |
| | Content source | Pre-encoded videos & real-time content |
| | Output rendering devices | GPUs |

| | | |
|-------------|----------------------------|---|
| | Single / Multiple machines | Multiple (for Deep Space 8K) and single (for HMDs) |
| Audio | Input | 5.1 |
| | Output | 5.1 |
| Interaction | Operations | Play, pause, seek, zoom, re-centering, show & hide mini-map, jump to next video |
| | Device | XBOX controller |

4.5.3 Encoder Settings

| | | |
|-----------|---------------|---|
| Video | Encoder | Spin Enc Offline (spinnffmpeg) |
| | Coding scheme | HEVC Main 10 Profile (4:2:0 10-bit) |
| | GoP size | 16 frames (hierarchical) |
| | Intra period | 64 frames |
| | Rate control | CBR |
| | Bitrate | 250 Mbps |
| Audio | Encoder | Audio is not encoded (uncompressed format used) |
| | Rate control | CBR |
| | Bitrate | 4608 Kbps |
| Container | | HEVC, WAV (PCM) |

4.5.4 Playback Settings

| | |
|--------------|--|
| Media player | Unity3D player (using native C++ Immersify plugin) |
| OS | Windows 7 SP1 64 Bit |
| CPU | 2x Intel Xeon CPU ES-2687W (2x 16 cores) v3 @ 3.30 Ghz |
| GPU | 4x Nvidia Quadro P6000 |
| Memory | 64 GB (8x 8 GB, DDR4 2400) |

4.6 Status

4.6.1 Expected Results

The expected results for Task 5.3 is:

“Guidelines for encoding and decoding immersive content, as well as media production workflows experiences. As a best practice reference, the information will also be published on the project website.”

4.6.2 Obtained Results

In Task 5.3 we created a list of recommendations for professionals to properly encode and playback immersive content including: for 8K 2D, 8K 3D (stereoscopic), high-resolution 360° video, point cloud video renderings, and interactive video playback using game engines. The guidelines describe the configuration of the encoder including the recommended bitrate that has been estimated using subjective testing, as well as the configuration of the media player and the corresponding minimum systems specifications required for the playback of the different immersive media target in the project.

5 Content Preparation Guidelines

In order to share experiences on producing immersive content we created guidelines for content creators on how to use the video compression tools for delivering and exhibiting immersive content. As a best practice reference, the information is also published on the project website. This will help to increase the adoption of state-of-the-art IT technologies by the media industry.

In this section we present a brief summary of each of the content preparation guidelines. The full text is available on the immersify website.

5.1 Point Cloud

This document describes how to scan 3D images with laser scanners into a point cloud and briefly presents the available formats and software. In addition, it describes the possibilities of editing and post-production of point clouds to produce high-resolution video in various formats: 8K and more, 3D, 360° along with a discussion of projection types such as ERP or CMP. It also discusses the preparation and encoding of the final image using software produced by the Immersify project. The document contains practical experience gained during the scanning of the cathedral in Poznan.

<https://immersify.eu/guidelines-reports/point-cloud/>

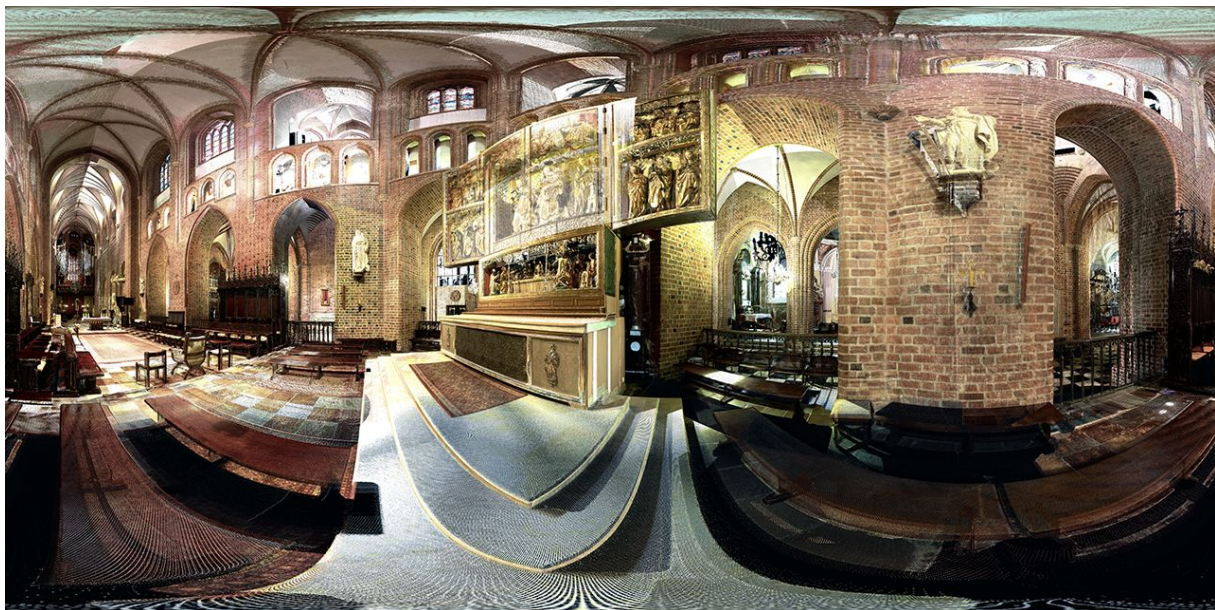


Figure 14. 360° version of *Poznan Cathedral* rendered from point clouds

5.2 8K 2D/3D Filming

This document presents methods of recording 8K content with high resolution cameras and photo cameras (timelapse). It presents methods of using 8K camera sets to obtain stereoscopic images using specialized 3D rigs. Practical examples present issues related to recording, data formats,

storage, post-production, color-grading, editing and rendering the final video. It also discusses methods and formats for preparation for encoding with codecs developed in the project.

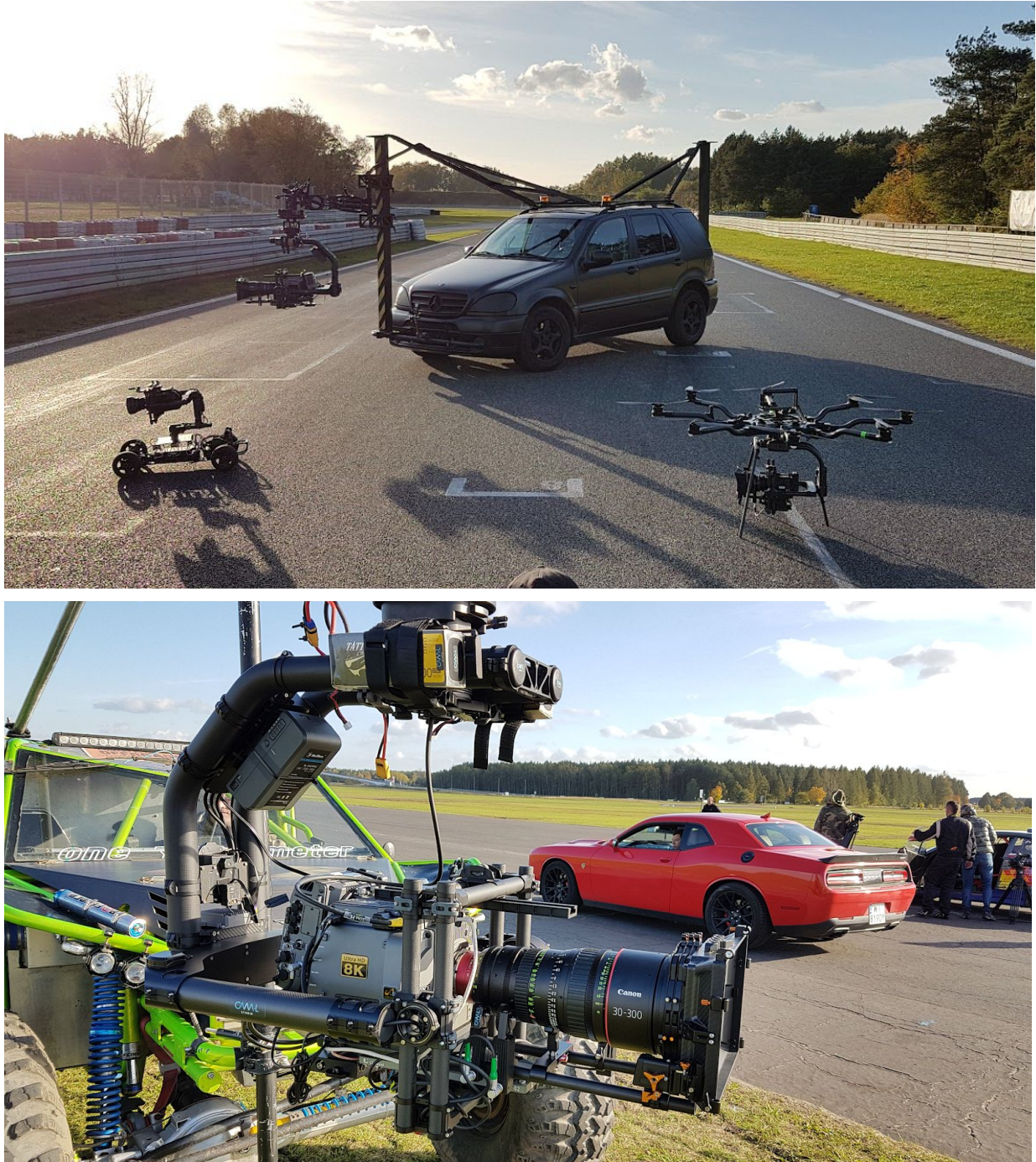


Figure 15. Production of *Follow Car* using 8K cameras

<https://immersify.eu/guidelines-reports/8k-2d-3d-filming/>

5.3 Ambisonic Sound Production

This document contains a description of the practical recording and production of ambisonic sound based on the experience gained in the Immersify project. The document presents practical methods

of ambisonic sound recording supported by examples of three realizations in different acoustic conditions and using different equipment. In addition, it discusses how to mix the sound using mostly free software. Finally, a practical example of a multi-speaker installation for ambience sound recording is presented. Each of the chapters is accompanied by practical advice and hints on how best to prepare for ambience recordings.



Figure 16. Ambisonics audio installation at PSNC

<https://immersify.eu/guidelines-reports/ambisonic-sound-production/>

5.4 Dome and Interactive Content

This document aims to explain how interactive content can be created for a dome or 360° environment. The focus is especially on how to create multi-user interactive narratives within a shared space. At the forefront is the use of game engines in the creation of dome experiences - how realtime software can be utilized to make user participation a main part of the experience. The document looks specifically at real-time production, immersion, interaction and multi-machine vs single-machine playback. It also shines a light on certain practices, such as creating specific camera setups, creating paths and coding the software with content creators in mind.

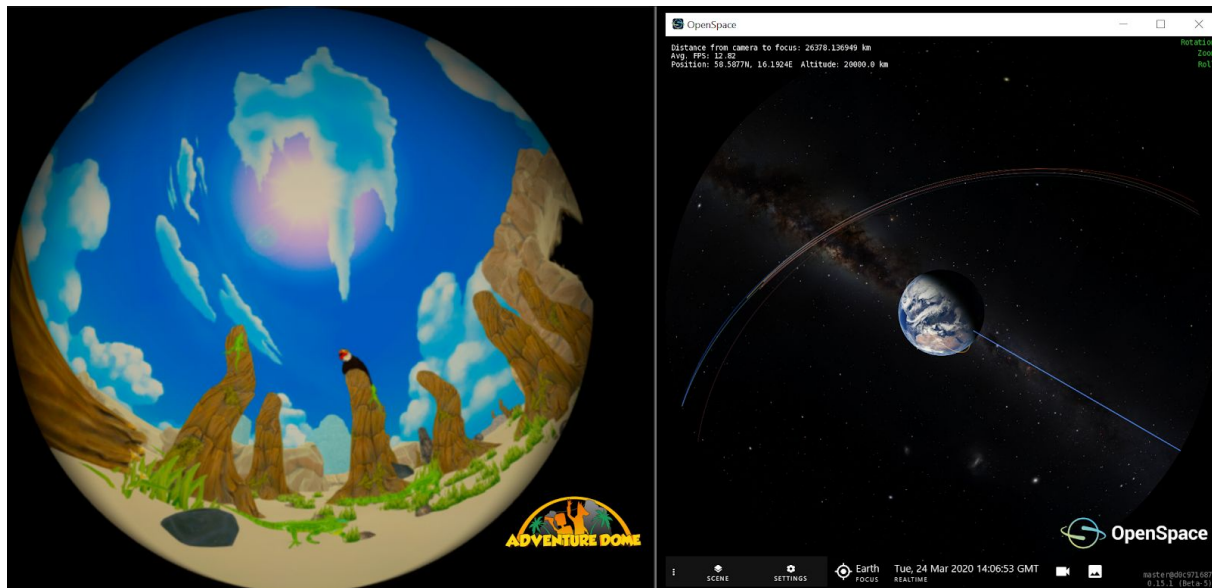


Figure 17. Fisheye camera views in used interactive applications

Figure above shows the two interactive projects/platforms created and/or extended to support the production of online and offline/video immersive for the dome (hence the fisheye camera) as well as other forms of immersive environments, such as Deep Space 8K or HMD headsets.

<https://immersify.eu/guidelines-reports/dome-and-interactive-content/>

5.5 Interactive Authoring, Deep Space DevKit

The Ars Electronica Deep Space 8K Unity Development Kit shall make it possible for developers all over the world to easily create Unity applications for the Deep Space 8K, a large-scale multiuser VR environment. This guideline gives an overview of the Deep Space 8K possibilities and explains the SDK in detail.

The Deep Space Dev Kit supports all available interfaces in the Deep Space and supports fast application development for this VR room. Supported hardware is:

- Pharos Laser Tracking System: This tracks the position of people, standing in the projection area.
- xBox Controller: Prepared to be easily used in Deep Space.
- Mobile Control: A special app, that can be filled dynamically with content that helps to control the content.
- VRPN: A well known protocol to simulate button presses or send the Deep Space mobile phone acceleration data.

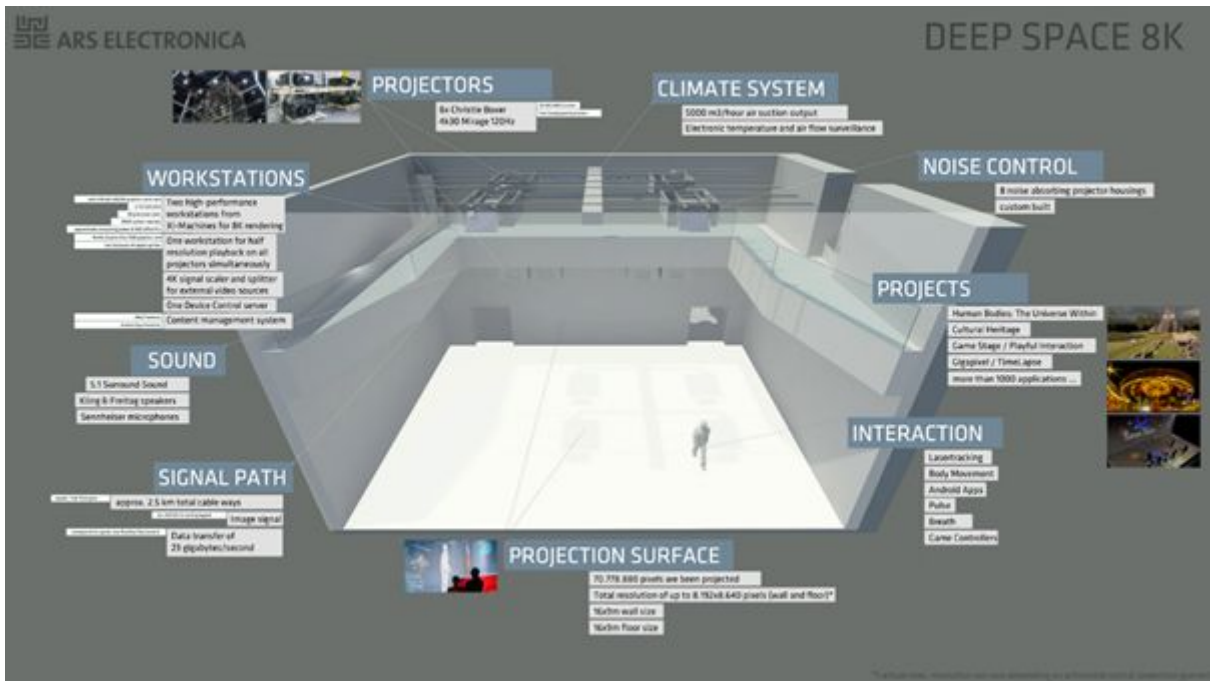


Figure 18. General overview of Deep Space 8K, summing up all important components

To enable a faster development, Deep Space DevKit comes with following commonly used components:

- Wall-Floor-Offaxis-Camera: A premade setup to combine the views for the wall and floor projection for an immersive experience.
- Command Line Config Manager: Enables the configuration of an application easily without having much programming effort.
- Networking: Synchronization of objects between wall and floor (e.g. the camera position) is already prepared and can easily be extended to fulfill other needs.

A complete documentation can be found here:

<https://immersify.eu/guidelines-reports/deep-space-development-kit/>

5.6 AI-Based Super Resolution Upscaling

Upscaling of digital images (or video frames) refers to increasing the resolution of the original images. This can be very useful when artists work with (older) low-resolution materials that need to be used in a high-resolution output file.

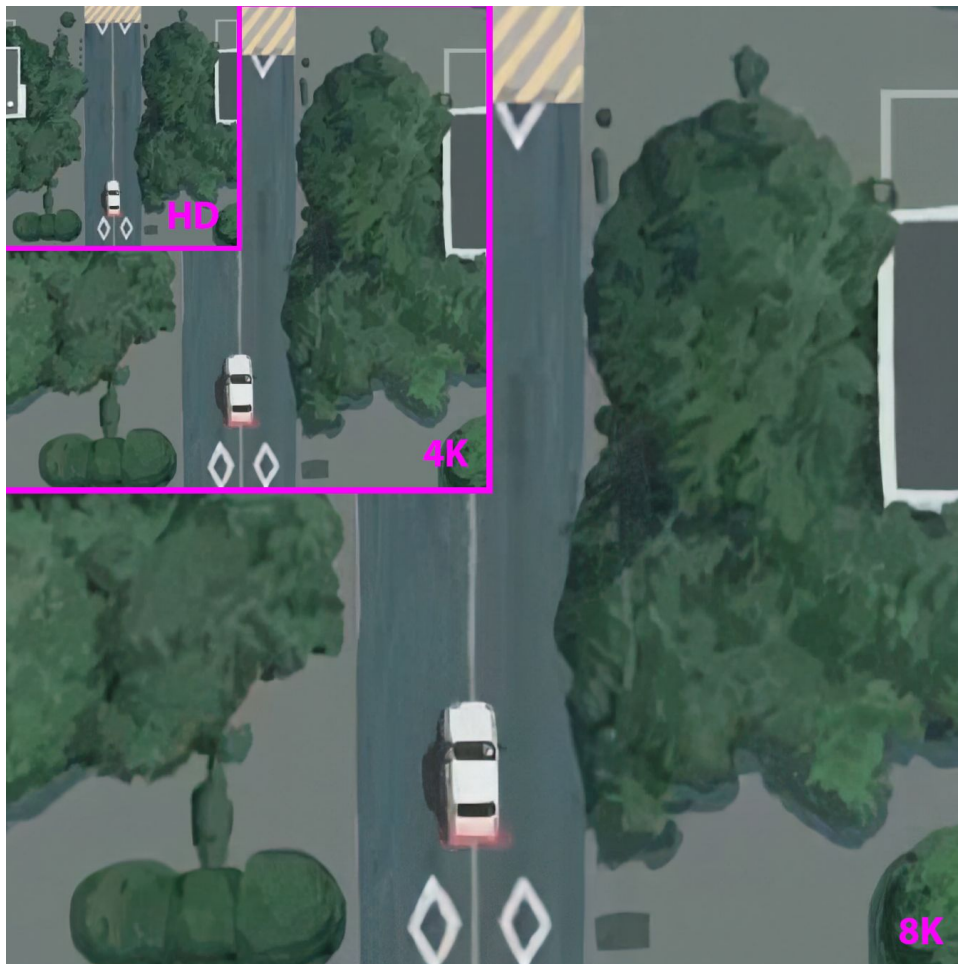


Figure 19. Comparing an original HD frame²⁰ to 4K and 8K upscaling via Topaz Video Enhance AI²¹

There are different approaches to upscaling. Some of them have already been integrated into commonly used image editing programs such as Adobe Photoshop or Aftereffect. Some others are available as standalone applications. In the context of this guide, we only focus on methods based on artificial intelligence and give an introduction to the employment of Deep Learning technology for upscaling video.

The target group of this guide are artists who have no previous experiences in the field of Deep Learning systems and want to use deep learning technology with as less effort as possible for upscaling their media content.

<https://immersify.eu/guidelines-reports/super-resolution-upscaling/>

²⁰ The example frame is taken from Sung Rok's animation "Scroll Down Journey"

²¹ <https://topazlabs.com/gigapixel-ai/>

5.7 Status

5.7.1 Expected Results

The expected results for Task 5.3 is:

“Guidelines for encoding and decoding immersive content, as well as media production workflows experiences. As a best practice reference, the information will also be published on the project website.”

5.7.2 Obtained Results

We created a total six guidelines for content creators that include our experiences on producing immersive media content using different technologies, such as 8K cameras, laser scanners, game engines, and ambisonics sound. Each use case requires a different production workflow that is described in detail on the project website:

<https://immersify.eu/guidelines-reports/>