|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| D:\usr\campos\TSB-Reference\Logos\ITU\sigleITU.gif | INTERNATIONAL TELECOMMUNICATION UNION  **TELECOMMUNICATION STANDARDIZATION SECTOR**  STUDY PERIOD 2017-2020 | | | SG12-C.n | |
| **STUDY GROUP 12** | |
| **Original: English** | |
| **Question(s):** | | Q13/12 | Geneva, 27-28 Feb 2018 | | |
| **CONTRIBUTION** | | | | | |
| **Source:** | | Huawei Technologies Co. Ltd. | | | |
| **Title:** | | The proposed baseline for G.360-VR | | | |
| **Purpose:** | | [Choose a purpose from the dropdown list] | | | |
| **Contact:** | | Rachel Huang Huawei Technologies Co. Ltd. China | | | Tel: +86 25 56623633 Fax:  E-mail: rachel.huang@huawei.com |
| **Contact:** | | Haoping Yu Futurewei Technologies Inc. USA | | | Tel: +1 317 965 9895 Fax:  E-mail: haoping.yu@huawei.com |
| **Contact:** | | Ashutosh Singla Technical University Ilmenau Germany | | | Tel: +49-3677-692757 Fax: +49-03677 69-1255 E-mail: [ashutosh.singla@tu-ilmenau.de](mailto:ashutosh.singla@tu-ilmenau.de) |
| **Contact:** | | Werner Robitza Deutsche Telekom AG  Germany | | | Tel: +49-3677-692757 Fax: +49-03677 69-1255 E-mail: werner.robitza@telekom.de |
| **Contact:** | | Alexander Raake Technical University Ilmenau Germany | | | Tel: +49-3677-692757 Fax: +49-03677 69-1255 E-mail: [alexander.raake@tu-ilmenau.de](mailto:alexander.raake@tu-ilmenau.de) |

|  |  |
| --- | --- |
| **Keywords:** | Insert keywords separated by semicolon (;) |
| **Abstract:** | This contribution proposes initial baseline text for G.QoE-VR |

**Introduction**

In the Q13/12 interim meeting in Krakow, some of the text below has already been discussed, but it was agreed to adopt it as a baseline text for G.VR-360. This proposal is for this purpose and it also contains the change marks compared to the version presented at that meeting (Contribution Huawei-3, Q13/12, Krakow).

**Proposal**

**CONTENTS**

| Page |
| --- |
| [1. Scope 3](#_Toc505961463)  [2 References 3](#_Toc505961464)  [3 Definitions 4](#_Toc505961465)  [3.1 Terms defined elsewhere 4](#_Toc505961466)  [3.2 Terms defined in this Recommendation 4](#_Toc505961467)  [4 Abbreviations and acronyms 4](#_Toc505961468)  [5 Conventions 4](#_Toc505961469)  [6 Framework 4](#_Toc505961470)  [7 Source stimuli 6](#_Toc505961471)  [7.1 Source signals recordings 6](#_Toc505961472)  [7.2 Video considerations 6](#_Toc505961473)  [7.3 Audio considerations 6](#_Toc505961474)  [7.4 Interaction considerations 6](#_Toc505961475)  [7.5 Duration of stimuli 7](#_Toc505961476)  [8 Test method, environment and subjects 7](#_Toc505961477)  [8.1 Test method 7](#_Toc505961478)  [8.2 Environment 7](#_Toc505961479)  [8.2.1 Controlled environment 7](#_Toc505961480)  [8.2.2 Public environment 8](#_Toc505961481)  [8.3 Subjects 8](#_Toc505961482)  [9 Experiment design 8](#_Toc505961483)  [10 Experiment implementation 8](#_Toc505961484)  [11 Data analysis 8](#_Toc505961485) |

Recommendation ITU-T <No.>

The Methodology of Quality of Experience (QoE) Evaluation for 360 Degree Virtual Reality Video

Summary

TBD

Keywords

Insert keywords separated by semicolon (;)

1. Scope

Virtual Reality (VR) is a new type of media different from the traditional video and audio media. It generates realistic images, sounds and other sensations that replicate a real or imaginary environment, and simulates a user's physical presence in this environment, by enabling the user to interact with this space and any objects depicted therein using specialized display screens or projectors and other devices. As one of the most important type of VR content, VR video, a.k.a., spherical/360-degree, had a fast development in the past few years. In contrast to the traditional video, 360-degree video can capture all 360 degrees of a scene, simultaneously. Typically, users view 360-degree videos on Head-Mounted Display (HMD) and can turn around to view the immersive 360-degree space from different angles. Coupled with the large field of view (FOV) presented by HMD, 360-degree video can provide a relatively immersive experience, which is significantly different from the experience provided by traditional video.

This recommendation describes methods to evaluate the sense of presence from viewing 360-degree video on an HMD. In general, this method utilizes a hierarchical design, where the entire evaluation process is divided into three abstraction layers. This Recommendation may be used to compare 360 degree VR devices performance in multiple environments, and to compare the quality impact of multiple 360 degree VR devices.

# 2 References

The following ITU-T Recommendations and other references contain provisions, which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published.

The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

[ITU-T H.ILE-SS] TD 11-WP3/16, January 2017 *Service scenario of immersive live experience (ILE)*.

[MPEG-I Part 1] N16918 “Working Draft 0.2 of Technical Report on Immersive Media”, MPEG 118, April 2017

[1] M. Slater and S. Wilbur, “A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments,” Presence: Teleoperators and Virtual Environments, vol. 6, no. 6, pp. 603-616, Dec 1997.

[2] J. J. Cummings and J. N. Bailenson, “How immersive is enough? A meta-analysis of the effect of immersive technology on user presence,” Media Psychology, vol. 19, no. 2, pp. 272-309, 2016.

[3] A. Singla, S. Fremerey, W. Robitza, P. Lebreton A. Raake, “Comparison of Subjective quality Evaluation for HEVC Encoded Omnidirectional Videos at Different Bit-rates for UHD and FHD”, in 25th ACM Multimedia (ACM MM) Thematic Workshops, October 2017.

# 3 Definitions

## 3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

TBD

## 3.2 Terms defined in this Recommendation

# 4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

|  |  |
| --- | --- |
| VR | Virtual Reality |
| HMD | Head mounted device |

# 5 Conventions

TBD

# 6 Framework

The framework for assessing the sense of presence discussed in this contribution is shown in Fig. 1. It focuses on the novel experience perceived by a user from the 360-degree video rendered on a VR HMD. In this evaluation framework, a hierarchical structure is devised with three layers, and each layer comprises modules designated for sensory cues of relevant parameters such as video, audio, latency, etc. The composition of the evaluation framework as well as the relationship between the layers is discussed in detail below.



Fig. 1. Evaluation framework for user’s experience quality (presence)

The concept of immersion and presence studied in [1] is adopted here, which states that:

1) immersion refers to the objective level of sensory fidelity a VR system provides;

2) presence refers to a user’s subjective psychological response to a VR system.

Considering the measurability of immersive features discussed in [2], two individual fidelity aspects from multi-modal sensory cues, namely the visual fidelity and acoustic fidelity, are selected to represent the quality of immersive media, and an interactive quality, namely sensory modality matching is used to evaluate the interactive feature of a VR system. These features are measured by the evaluation platform, as shown in Fig. 1, according to the following definition:

1) Visual fidelity ---- visual fidelity refers to how close the system’s visual output, i.e., the virtual scene provided by the 360-degree video and HMD, is to real-world visual stimuli.

2) Acoustic fidelity ---- like the visual fidelity, this feature represents how close the system’s aural out-put, i.e., the sound provided by the stereo or spatial audio of a VR system, is to real-world aural stimuli.

3) Sensory modality matching ---- this feature refers to the matching degree, during head movements, between the sensory modalities, such as proprioception, visual and auditory perceptions.

It should be noteworthy that this framework is only focused on the presence referring to a sense of “being there”, without considering other aspects, such as social, self, etc., of the sense of presence introduced in [3].

# 7 Source stimuli

In order to evaluate quality in various circumstances, the content should cover a wide range of stimuli. The stimuli should be selected according to the goal of the test and recorded on a digital storage system. When the experimenter is interested in comparing results from different laboratories, it is necessary to use a common set of source stimuli to eliminate a further source of variation.

The selection of the test material should be motivated by the experimental question addressed in the study. For example, the content of the test stimuli should be representative of the full variety of programmes delivered by the service under study (sport, drama, film, speech, music, etc.).

## 7.1 Source signals recordings

The source signal provides the reference stimuli and the input for the system under test.

The quality of the reference stimuli should be as high as possible. As a guideline, the video signal should be recorded in uncompressed multimedia files using one of the following two formats: YUV (4:2:2 or 4:4:4 sampling) or RGB (24 or 32 bits).

## 7.2 Video considerations

The selection of source video is an important issue. Both monoscopic and stereoscopic contents should be allowed. The set of test scenes should span the full range of spatial and temporal information of interest to users of the devices under test. The quality of the SRCs should also be as equal as possible. It is suggested to use 4K or higher resolution video sequence to avoid the situation that low resolution of the original VR content enlarged on VR near-eye display causes the absence of the presence experience.

## 7.3 Audio considerations

[Editor’s note]: This section needs more investigation.

When testing the quality and fidelity of audio in 360 degree VR sequence, the speech need not be in a language understood by all subjects. Both stereo and spatial audio can be used.

## 7.4 Interaction considerations

360 degree VR application is weak-interactive VR, where users passively experience pre-filmed content in a virtual environment. Users alter viewpoints by swivelling but cannot engage in substantial interaction with the virtual environment.

Accordingly, the interaction experience of 360 degree VR is mainly reflected in motion-to-photons latency (MTP). The set of test sequences should span a range of different MTP time to users of the devices under test.

.

## 7.5 Duration of stimuli

The methods in this Recommendation are intended for stimuli that range from 10 seconds to 5 minutes in duration. For subjects who participate in the test at their first time, it is suggested to use long sequences, e.g., from 30 seconds to 5 minutes, to allow them to have enough time to accommodate themselves to the virtual environment.

As 10 seconds to 20 seconds may be quite short, it is suggested to use methods, e.g., M-ACR defined in next sections.

# 8 Test method, environment and subjects

## 8.1 Test method

## 8.1.1 Absolute Category Rating

This Recommendation uses the absolute category rating (ACR) method, which is a category judgement where the test stimuli are presented one at a time and are rated independently on a category scale. The subject observes one stimulus and then has time to rate that stimulus.

The ACR method uses the following five-level rating scale:

5 Excellent

4 Good

3 Fair

2 Poor

1 Bad

The numbers may optionally be displayed on the scale.

## 8.1.2 Modified Absolute Category Rating

M-ACR in [3] is also proposed for evaluating the 360° videos in an HMD based displays. Presentation of one stimulus is shown in Fig. 1. The video under test was shown twice and a mid-grey screen was shown for 6 s in between the test videos. After the second test video was shown, subjects were asked to rate the quality of the video on a 5-point scale (1: Bad, 2: Poor, 3: Fair, 4: Good, 5: Excellent).

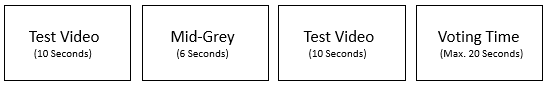


Figure 2. Presentation of one stimulus in M-ACR method [2].More experiment information and results regarding M-ACR method investigation can be found in Appendix I.

## 8.1.2 Double Stimulus Impairment Scale Test Method

In DSIS test method, first, a reference video was shown followed by the video under test and in between reference video and video under test a mid-grey screen was shown for 3 seconds. After the video under test is shown, the user is asked to rate compression artifacts and impairments on a 5-point scale (1: Very annoying, 2: Annoying, 3: Slightly annoying, 4: Perceptible but not annoying, 5: Imperceptible). The rating scale was projected at 4 different azimuth angles. The presentation of one stimulus for DSIS test method is shown in Fig.3.

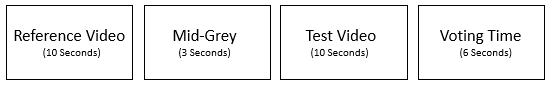


Figure 3. Presentation of one stimulus in DSIS method [3].

[Editor’s note]: Other methods are considered later.

## 8.2 Environment

The environment is not rigorously constrained within this Recommendation. Exceptions include cases where the experiment is designed to investigate the impact of a particular part of the environment on MOS (e.g., the impact of HMD type on MOS).

This Recommendation allows two options for the environment in which the subjective experiment takes place:

• a controlled environment

• a public environment.

The environment must be documented.

## 8.2.1 Controlled environment

A controlled environment should represent a non-distracting environment where a person would reasonably use the device under test. In this Recommendation, the test should be carried out in a noise isolation environment, as noise is the major environmental factor that would affect the test. Environmental factors other than noise are shielded out by the HMD and thus do not impact the subjects.

## 8.2.2 Public environment

A public environment should represent a distracting environment where a person would reasonably use the device under test. In this Recommendation, the environment should be allowed to have noise or other factors that could influence the user

## 8.3 Subjects

At least 24 subjects must be used for experiments conducted in a controlled environment. This means that after subject screening, every stimulus must be rated by at least 24 subjects.

# 9 Experiment design consideration

When stimuli with intermittent impairments are included in an experiment, care must be taken to ensure that the impairment can be perceived within the artificial context of the subjective quality experiment. The first one second and the last one second of each stimulus should not contain freezing, rebuffering events, and other intermittent impairments.

Examples of intermittent impairments include but are not limited to:

• pause, then playout resumes with no loss of content (e.g., “stalling” or “initial playout delay”);

• pause followed by a skip forward in time (e.g., transmission error causes temporary loss of signal, and system maintains a constant delay, or “freezing with skipping”);

• skip forward in time (e.g., buffer overflow);

• audiovisual synchronization errors (e.g., may only be perceptible within a small portion of the stimuli)

• packet loss with brief impact.

• MTP latency with no loss of content.

TBD.

For video-only experiments, the missing audio should not be considered as the impairment, and vice versa.

# 10 Experiment implementation

The experiment implementation follows the five stages defined in [P.913]:

1. Informed consent;
2. Pre-screening of subjects
3. Instructions and training
4. Voting sessions
5. Questionnaire and/or interview

# 11 Data analysis

After all subjects are run through an experiment, the ratings for each clip are averaged to compute either a mean opinion score (MOS)

**Appendix I**

**(**This appendix does not form an intergral part of this Recommendation**)**

This section makes an attempt to compare the performance of DSIS and M-ACR test methods on the basis of reliability and simulator sickness. In the following sections, we are providing a detailed description of the experimental setup, evaluation methodologies and experimental results.

# Experimental Setup

## *Description of the Test Dataset*

The dataset used in these two subjective test was downloaded from [4] [5] in two different resolutions 4K and 8K. Due to the hardware limitation, 8K sequences were down-sampled to 4K and FHD

Table I.1. Description of the Test Sequences

|  |  |  |  |
| --- | --- | --- | --- |
| Content No. | Content Name | UHD/4K | FHD |
| Bit-rate (Mbps) | Bit-rate (Mbps) |
| 1 | Driving in  Country | 1 | 0.5 |
| 2 | 1.0 |
| 4 | 1.5 |
| 8 | 3.5 |
| 15 | 5.0 |
| 2 | PoleVault\_le | 1 | 0.5 |
| 2 | 1.0 |
| 4 | 1.5 |
| 8 | 3.5 |
| 15 | 5.0 |
| 3 | Gaslamp | 1 | 0.5 |
| 2 | 1.0 |
| 4 | 1.5 |
| 8 | 3.5 |
| 15 | 5.0 |
| 4 | Harbor | 1 | 0.5 |
| 2 | 1.0 |
| 4 | 1.5 |
| 8 | 3.5 |
| 15 | 5.0 |
| 5 | KiteFlite | 1 | 0.5 |
| 2 | 1.0 |
| 4 | 1.5 |
| 8 | 3.5 |
| 15 | 5.0 |
| 6 | Trolley | 1 | 0.5 |
| 2 | 1.0 |
| 4 | 1.5 |
| 8 | 3.5 |
| 15 | 5.0 |

resolutions. The description of the dataset is given in Table 1 along with their bit-rates, content name and identifier. The downloaded sequences are in YUV 4:2:0 color space, 30fps, 8-bits per channel, in Equirectangular Projection (ERP) with a duration of 10 s. The source sequences were encoded using FFmpeg with *libx265* (HEVC/H.265) with the Video Buffering Verifier (VBV) method. *Libx265* has been chosen over HM and JEM reference software due to the higher encoding speed and better control over the bit-rates.

## *Test Environment*

The Oculus Rift (Consumer Version 1) was used for displaying the 360° videos to the users. The resolution of the device is 2160×1200 and the field of view is 110°. *Whirligig* player was used for playing the videos in an HMD. The Oculus Rift was connected to a desktop PC equipped with an NVIDIA GTX980 graphics card and an Intel Core i7 processor [2] [3].

## *Test method*

## *Modified-ACR Test Method*

This method is specified in Section 8.1.2. 30 users participated in the M-ACR study. Out of 30 users, 15 users were females and 15 were males with an average of 25.62 years.

## *Double Stimulus Impairment Scale Test Method*

27 subjects participated in the DSIS study. Out of 27, 16 were males and 11 were females with an average age of 25.07 years.

In both these subjective test methods, all the users were pre-screened for correct visual acuity and color blindness using Snellen (20/25) and Ishihara charts respectively. Instructions were given to the subjects not to consider stitching and ghosting artifacts while rating the videos.

# Experimental Results

Outlier detection was performed for checking the reliability of the users. Pearson correlation coefficient was computed between the MOS and the raw scores of each subject. A threshold of 0.75 was selected for detecting the outliers. Only one user is found to be an outlier in each of the test methods.

## *Comparisons of MOS between DSIS and M-ACR*

Figure I.2 shows the MOS along with the associated 95% Confidence Interval (CI) averaged over all the video sequences for the two test methods DSIS and M-ACR for the two different resolutions (4K and FHD). From the Fig. 3, it is clear that the subjects are able to find out the difference between the 4K and FHD resolutions in both these evaluation methodologies. M-ACR provides the higher MOS than DSIS test method for the lower bit-rates (1 and 2 Mbit/s) at 4K resolution. But at higher bit-rates (4, 8, 15 MBit/s), DSIS provides slightly better MOS as compared to M-ACR.

From both these evaluation methodologies, it is clear that the perceived quality at 15 Mbit/s is slightly higher than 8 Mbit/s for 4K resolution. Therefore, 360° videos can be transmitted at 8 Mbit/s at a marginal loss of quality but ensures almost 50% saving of the bandwidth.

FHD

4K

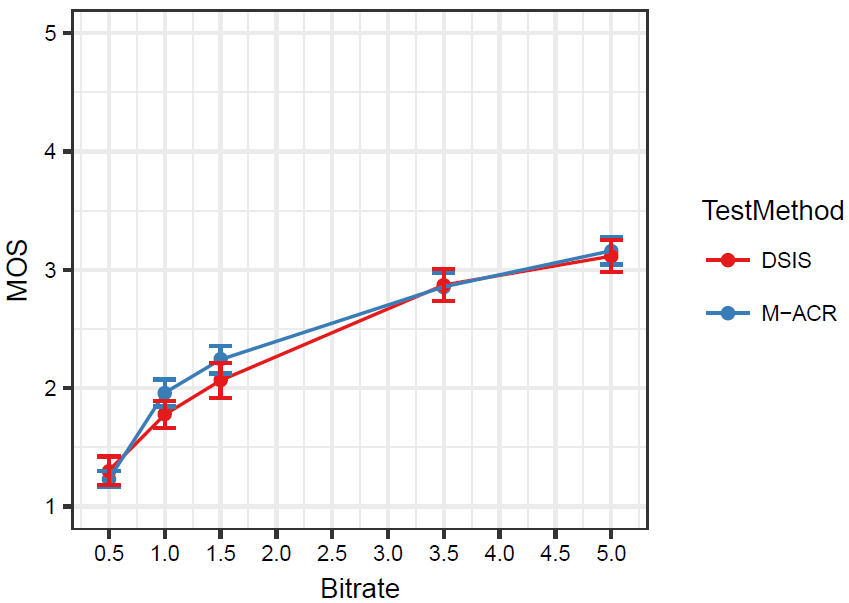
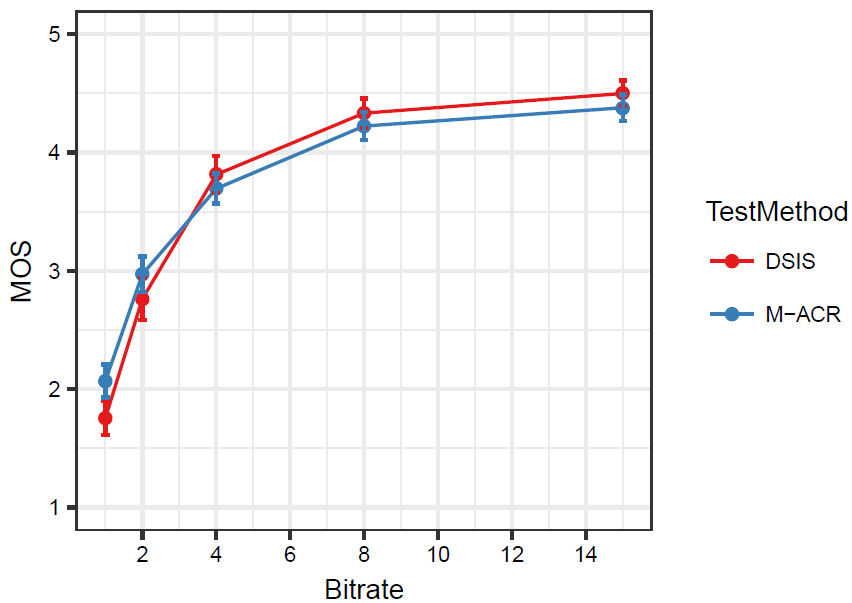
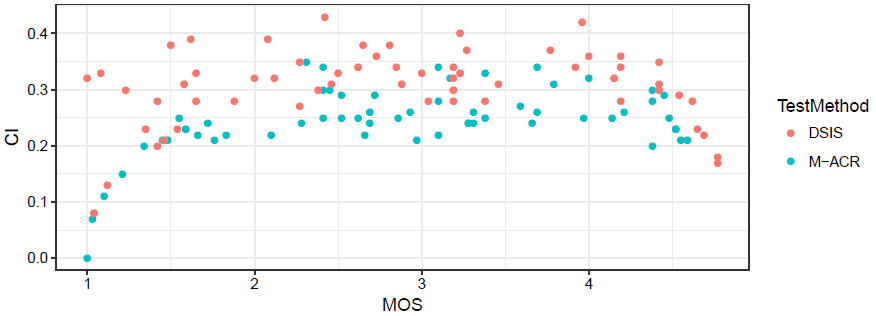


Figure I.2. Average MOS for different test methods at 4K and FHD resolutions [3].



. Figure I.3. CI vs MOS [3].

Figure I.3 shows the plot between CI vs MOS for DSIS and M-ACR test methods. This plot shows that DSIS provides a slightly higher range of MOS as compared to M-ACR, but it is also interesting to note that CI values are higher in DSIS method. DSIS evaluation methodology could be more useful for evaluating the minute and fine-grained details at 4K resolution due to its higher subjective resolution power.

We computed the statistical reliability based on [6] for check the reliability of these two evaluation methodologies. We calculated the MCInorm for M-ACR and DSIS based on the equation shown below:

MCI is the mean confidence interval averaged over all the video sequences, bit-rates and resolutions. Whereas MOS Range is the difference between the highest MOS and the lowest MOS for each of the test method. Table I.2 shows the MOS Range, MCI and MCInorm  for M-ACR and DSIS test methods. Results indicate that M-ACR is statistically more reliable as compared to DSIS.

Table I.2. Description of the Test Sequences [2]

|  |  |  |
| --- | --- | --- |
|  | M-ACR | DSIS |
| MCI | 0.118 | 0.136 |
| MOS Range | 3.149 | 3.198 |
| MCInorm | 0.037 | 0.042 |

## *Simulator Sickness*

Users may experience the symptoms of simulator sickness while evaluating the 360° videos in an HMD. Therefore, assessing the simulator sickness should be done in order to find out in which evaluation methodology users are more prone to simulator sickness. Assessment of simulator sickness was done by using the simulator sickness questionnaire based on [7]. Figure I.4 shows the simulator sickness scores for DSIS and M-ACR test methods. We have computed the simulator sickness scores for session #0 for DSIS method which were collected just after the pre-screening. The result shows that users are more prone to simulator sickness while evaluating the 360° videos using DSIS evaluation methodology except in session #1. The possible reason could be the changing of resolution and bit-rate between the reference and video under test in DSIS test method.

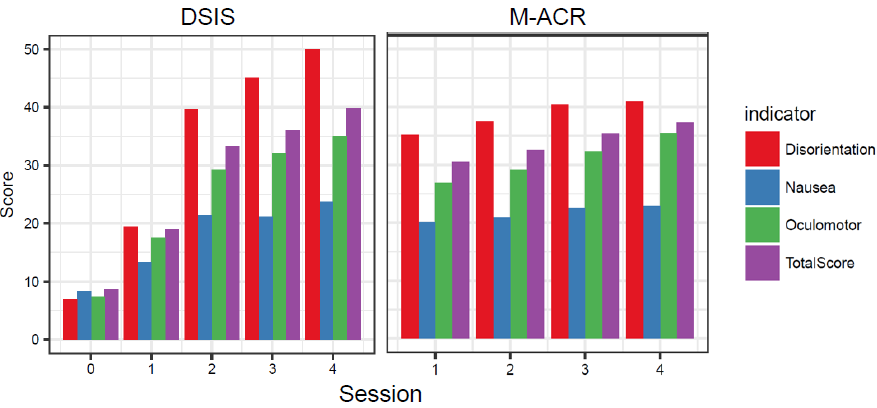


Figure I.4. Simulator Sickness Scores for all the test sessions for DSIS and M-ACR test methods [3].

# References

1. A. Singla, S. Fremerey, W. Robitza, A. Raake, “Measuring and comparing QoE and simulator sickness of omnidirectional videos in different head mounted displays”, in 9th International Conference on Quality of Multimedia Experience (QoMEX), May 2017.
2. A. Singla, S. Fremerey, W. Robitza, P. Lebreton A. Raake, “Comparison of Subjective quality Evaluation for HEVC Encoded Omnidirectional Videos at Different Bit-rates for UHD and FHD”, in 25th ACM Multimedia (ACM MM) Thematic Workshops, October 2017.
3. A. Singla, W. Robitza, A. Raake, “Comparison of subjective quality evaluation methods for omnidirectional videos with DSIS and Modified ACR”, Human Vision Electronic Imaging, January 2018.
4. E. Asbun, Y. He, Y. He, “AHG8: InterDigital Test Sequences for Virtual Reality Video Coding”, Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, [JVET-](http://phenix.int-evry.fr/jvet/doc_end_user/current_document.php?id=2736)D0039, 4th Meeting, Oct. 2016.
5. J. Boycee, E. Alshina, A. Abbas “JVET common test conditions and evaluation procedures for 360° video”, Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, [JVET-](http://phenix.int-evry.fr/jvet/doc_end_user/current_document.php?id=2736)D1030, 4th Meeting, Oct. 2016.
6. Tominaga, Toshiko, et al, “Performance comparisons of subjective quality assessment methods for mobile video”, in second IEEE international workshop on Quality of multimedia experience (QoMEX), 2010.
7. R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal, “Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness,” The International Journal of Aviation Psychology, 1993.