Comparison of Different Subjective Test Methods for HEVC Encoded Omnidirectional Videos

A. Singla, S. Fremerey, F. Hofmeyer, W. Robitza, Alexander Raake
Audiovisual Technology Group
Technical University Ilmenau Germany

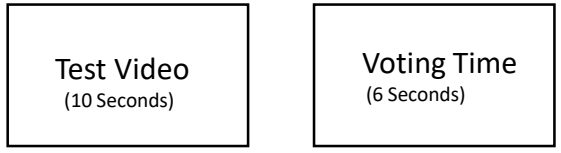
Introduction/Motivation

- An important feature of media streaming : Quality adaptation of 360° videos
- Resolution limitation of HMDs: Utilization of network resources efficiently
- What is the optimal bit-rate for watching the 360°video with an HMD?
  - Comparing DSIS, ACR and Modified-ACR scale
- Does simulator sickness change between test sessions?

A. Singla et. al., “Comparison of Subjective Quality Evaluation for HEVC Encoded Omnidirectional Videos at Different Bit-rates for UHD and FHD Resolution”, Thematic Workshops of ACM MM, October 2017

Video Quality Test Methods

ACR



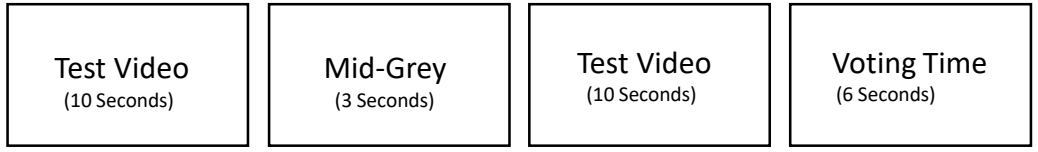
Presentation of One Stimulus in ACR

5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

ACR Scale

29 Participants
-11 Females
-18 Males
- Avg. age = 26.06

Modified-ACR



Presentation of One Stimulus in M-ACR

5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

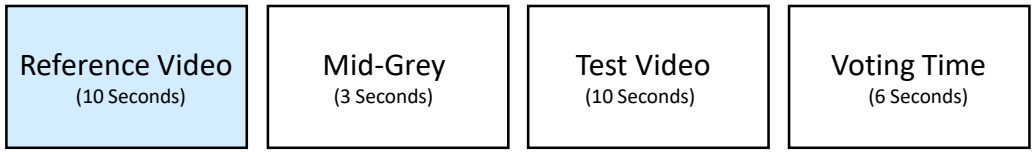
M-ACR Scale

30 Participants
-14 Females
-16 Males
- Avg. age = 25.6

A. Singla et. al., “Comparison of Subjective Quality Evaluation for HEVC Encoded Omnidirectional Videos at Different Bit-rates for UHD and FHD Resolution”, Thematic Workshops of ACM MM, October 2017

Video Quality Test Methods

DSIS



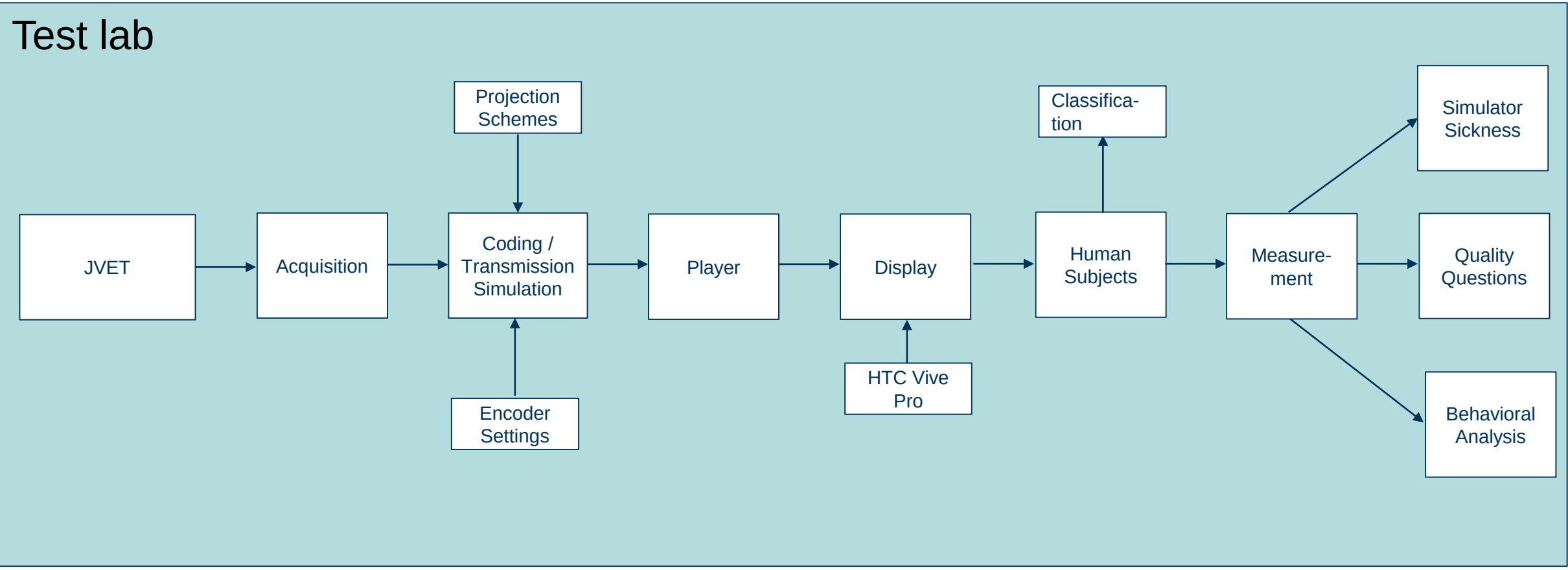
Presentation of One Stimulus in DSIS

5	Imperceptible
4	Perceptible, but not annoying
3	Slightly annoying
2	Annoying
1	Very annoying

´DSIS Scale

28 Participants
- 9 Females
- 19 Males
- Avg. age = 24.96

360° Video Test Framework TU Ilmenau



Dataset (8K, 6K, and 4K) – 10s videos

HTC Vive Pro
-2880×1600
-110° FOV
- Whirligig player

Content 1: Gaslamp

Content 2: harbor

Content 3: KiteFlite

Conten4: SkateboardInLot

Content 5: Trolley

Test Session

Total Duration (100 minutes)

Pre-screening, Training (10 minutes)	Pause (10 minutes)	Test Session 1 (11 minutes)	Pause (11 minutes)	Test Session 2 (11 minutes)	Pause (11 minutes)	Test Session 3 (11 minutes)	Pause (11 minutes)	Test Session 4 (11 minutes)

Example of One Test Session for DSIS and M-ACR

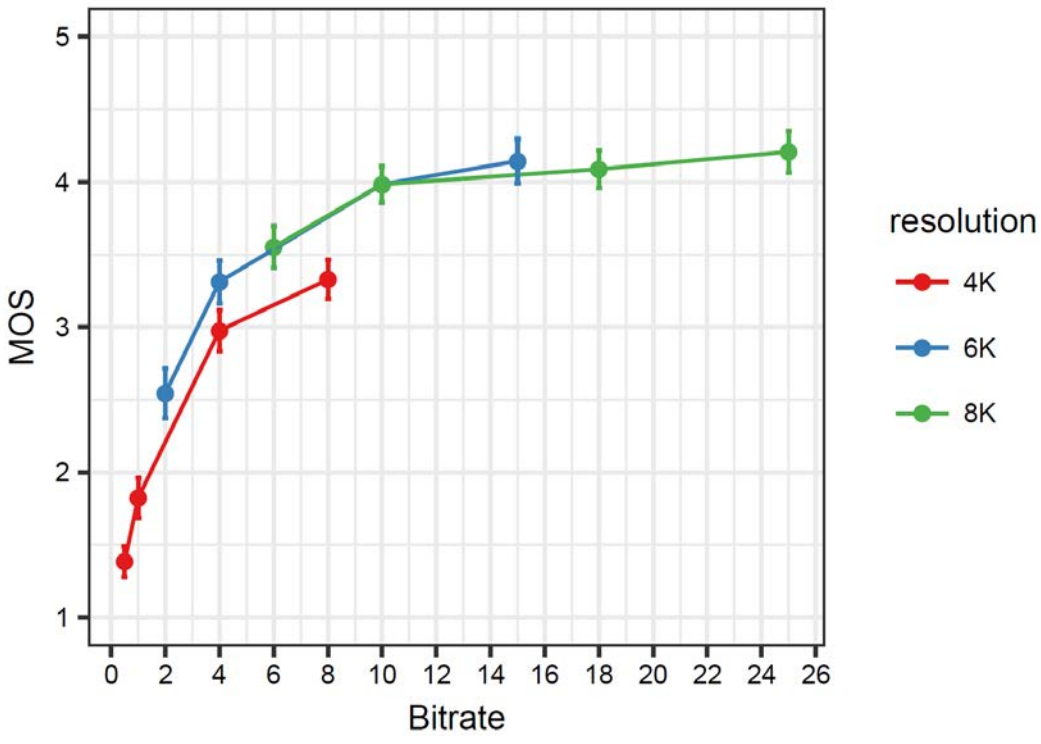
Total Duration (70 minutes)

Pre-screening, Training (10 minutes)	Pause (10 minutes)	Test Session 1 (7 minutes)	Pause (7 minutes)	Test Session 2 (7 minutes)	Pause (7 minutes)	Test Session 3 (7 minutes)	Pause (7 minutes)	Test Session 4 (7 minutes)

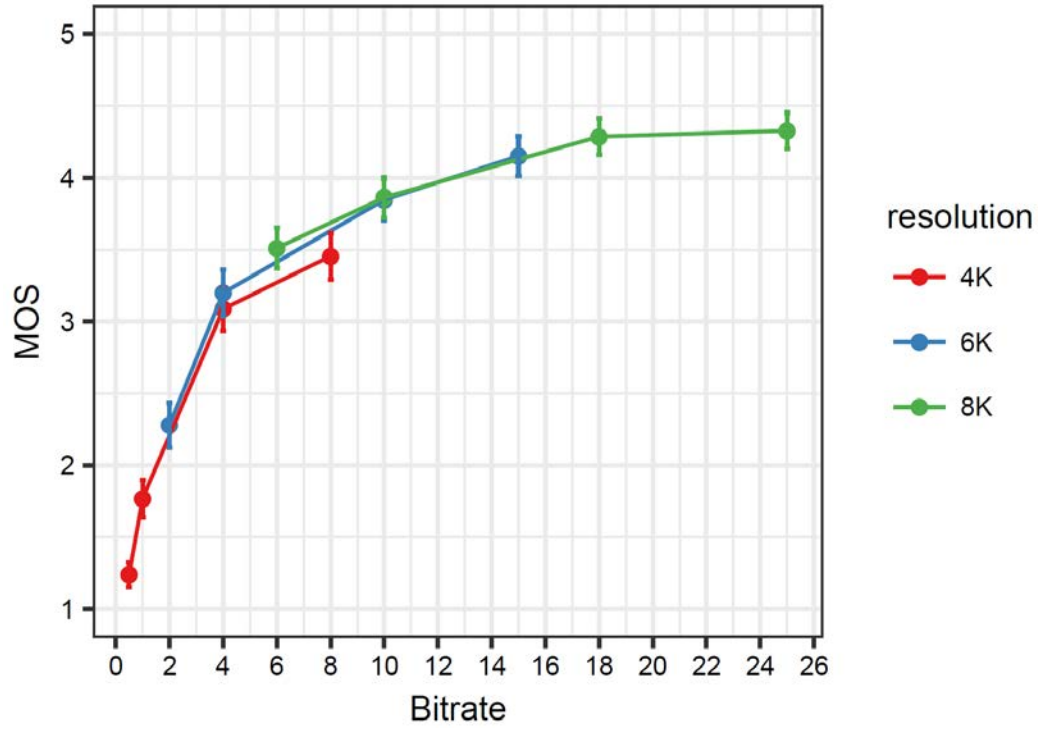
Example of One Test Session for ACR

Experimental results: Video Quality

ACR

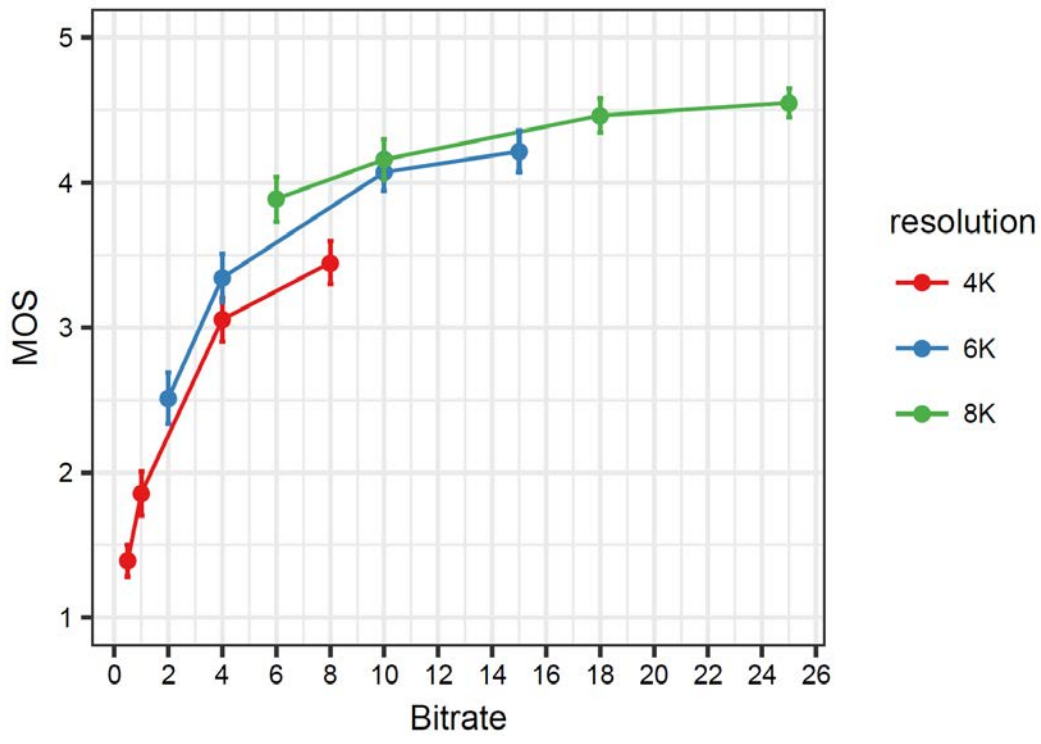


M – ACR

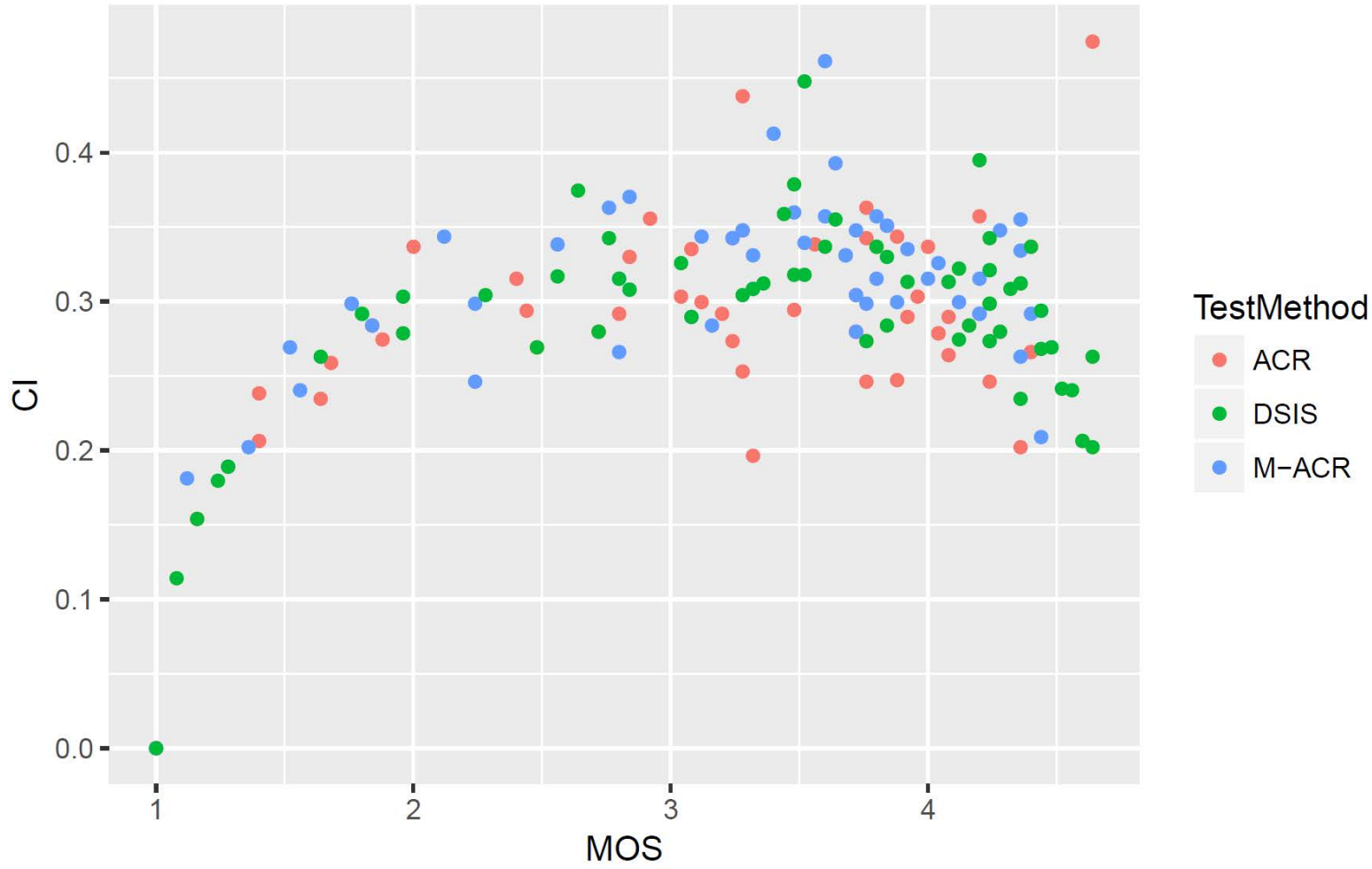


Experimental results: Video Quality

DSIS



Experimental results: CI vs MOS



CI : Confidence Interval

Experimental results: Statistical Reliability*

- MCI: Mean Confidence Interval
- MOS Range: Absolute difference between the highest and lowest MOS for each test method

$$MCI_{norm} = \frac{MCI}{MOS\ Range}$$

Table 1. MCI, MOS Range and MCI_{norm} for ACR, M – ACR and DSIS test methods

	ACR	M – ACR	DSIS
MCI	0.1382	0.1392	0.1405
MOS Range	2.824	3.088	3.16
MCInorm	0.0489	0.0450	0.0444

*Tominaga, et al. ”Performance comparisons of subjective quality assessment methods for mobile video”, in second IEEE international workshop on Quality of multimedia experience (QoMEX), 2010.

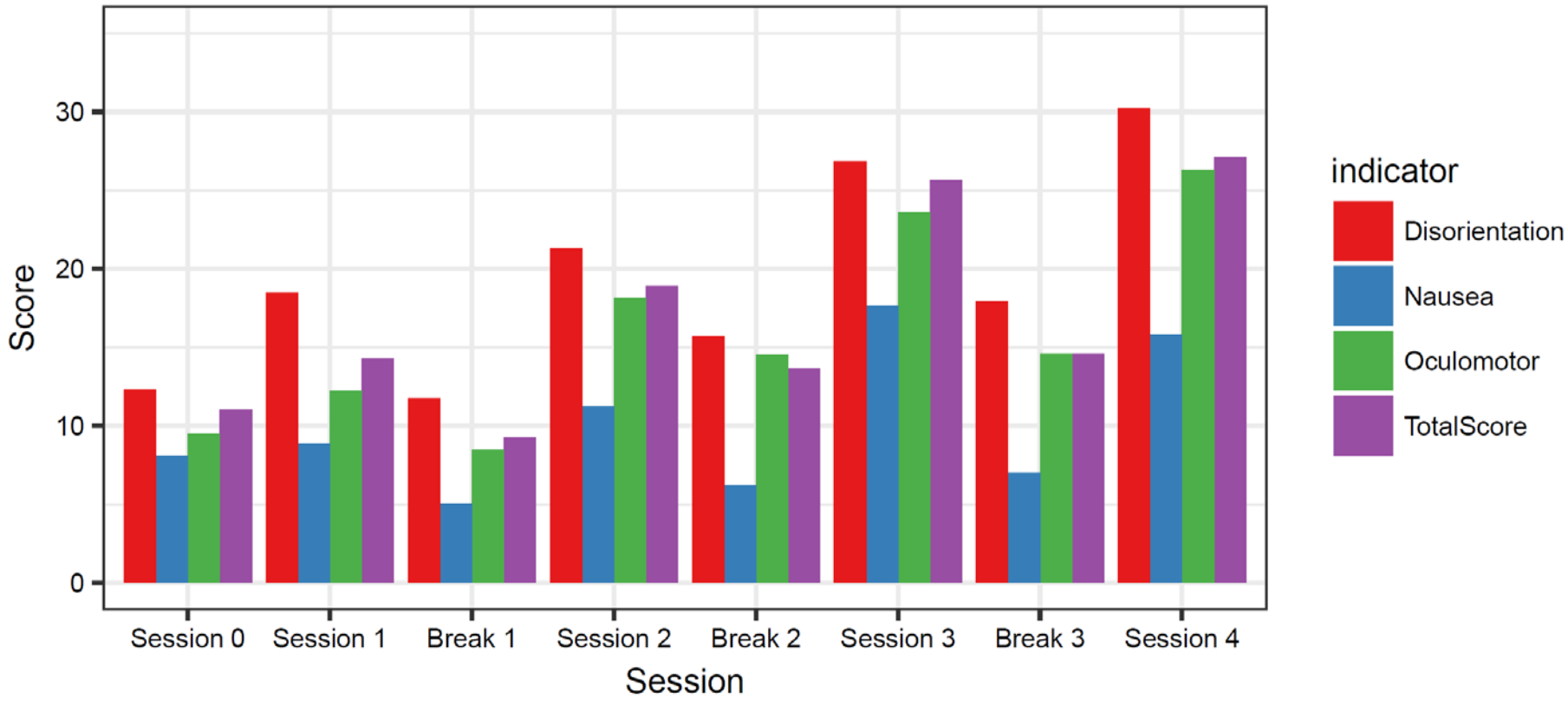
Symptoms of Simulator Sickness

Measurement of Simulator Sickness

- Measurement
  - 16 Questions
    - N (Nausea), O (Oculomotor), and D (Disorientation)
  - Nausea/Headache
    - None, Slightly, Moderate, Severe
  - 4-point scale is used
    - 0,1, 2, and 3
  - Total Score
    - ([N] + [O] + [D]) * 3.74

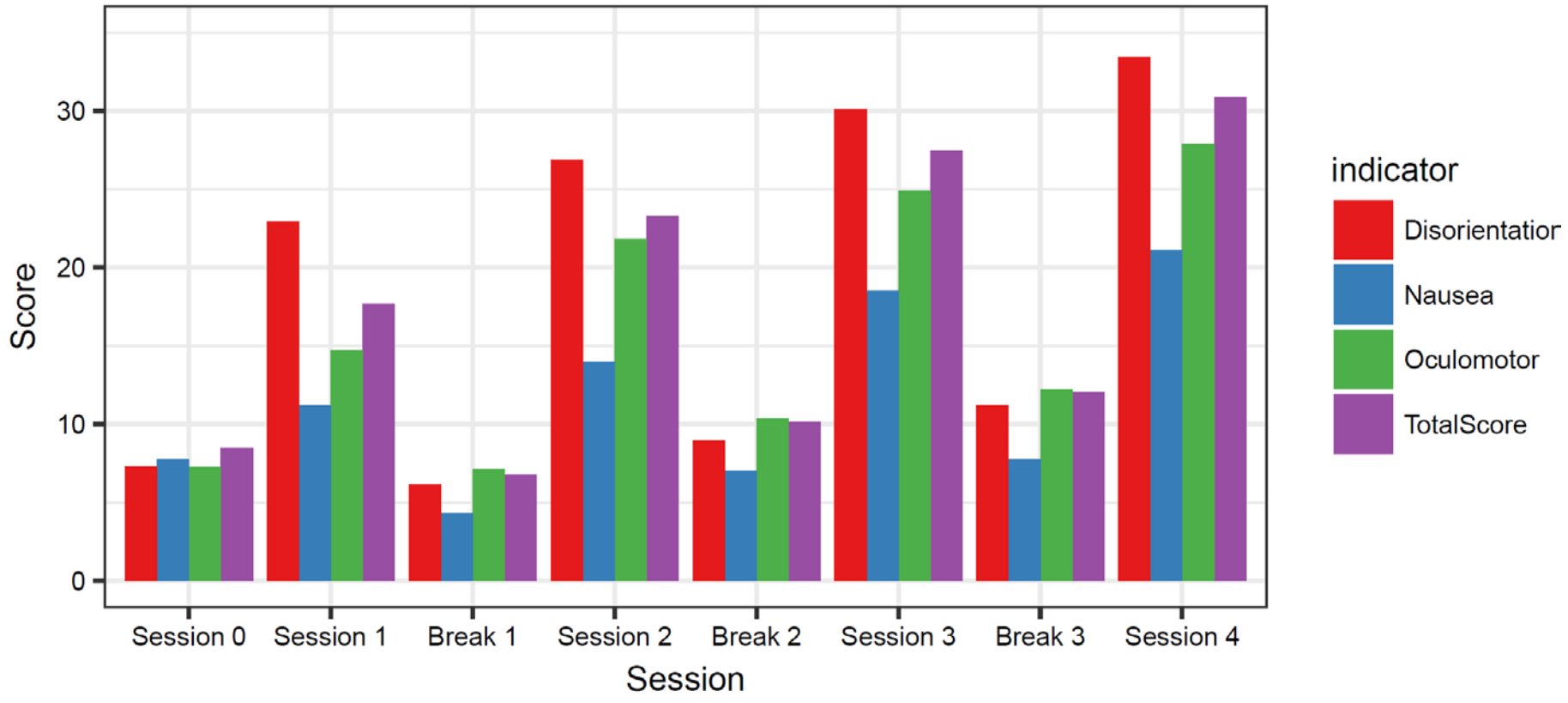
Experimental results: Simulator Sickness

ACR



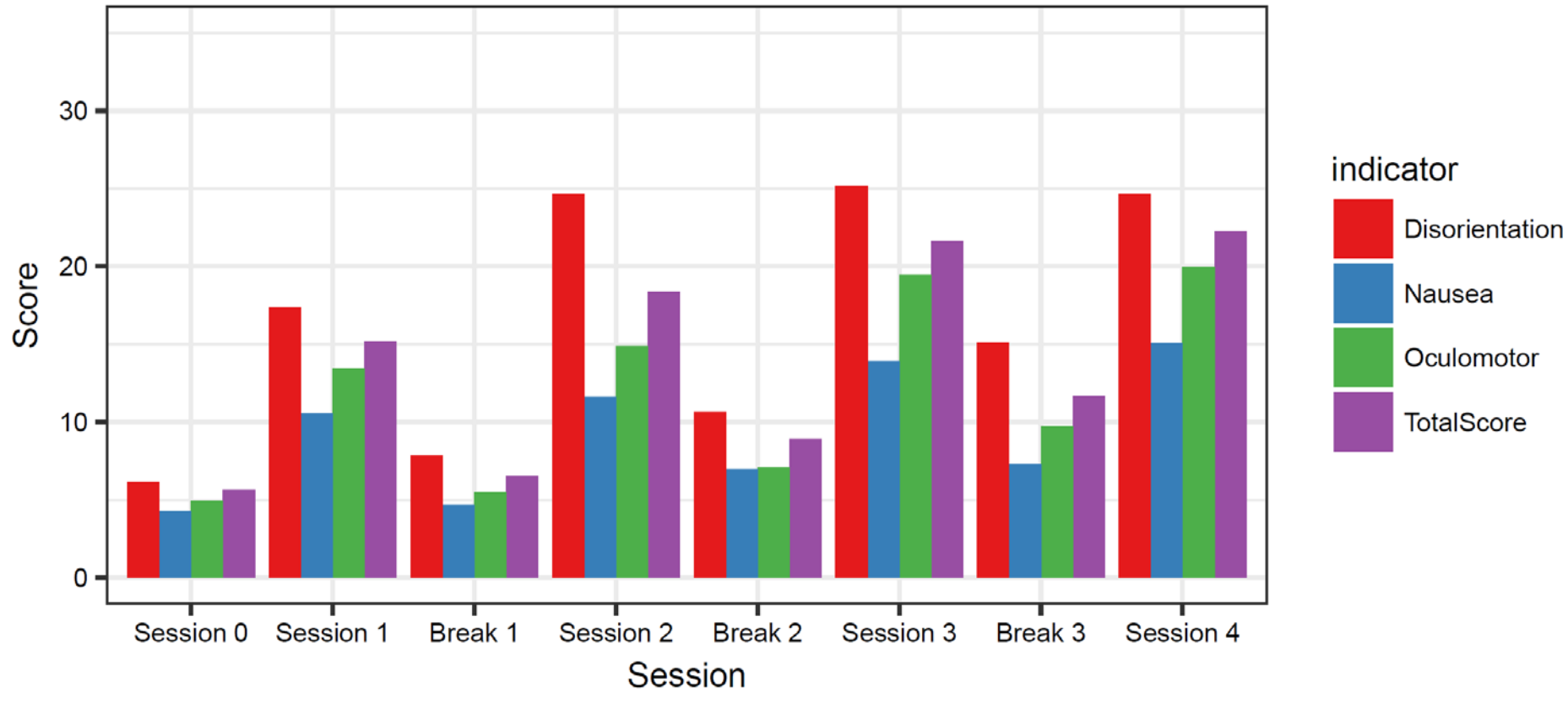
Experimental results: Simulator Sickness

M – ACR



Experimental results: Simulator Sickness

DSIS



Conclusion

- Video Quality
  - **6K provides better perceived quality as compared to 4K resolution**
  - 25 Mbps (8K) provides almost same perceived quality as 15 Mbps (6K)
  - 6K provides similar perceived quality as compared to 8K resolution
  - **DSIS statistically seen more reliable than ACR and M – ACR**
  - Very high correlation between test methods
    - ACR – M-ACR (Pearson Correlation coefficient = 0.95)
    - ACR – DSIS (Pearson Correlation coefficient = 0.93)
    - DSIS – M-ACR (Pearson Correlation coefficient = 0.97)

- Simulator Sickness
  - Simulator sickness scores increase with time
  - **Breaks help in reducing the simulator sickness scores**
  - Subjects are least prone to simulator sickness when evaluating 360°videos with DSIS

Are people pixel-peeping 360° videos?

-

Presented at the special session on QoE for immersive media at HVEI 2019)

Stephan Fremerey*, Rachel Huang**, Alexander Raake*
* *Audiovisual Technology Group, TU Ilmenau*
** *Huawei Technologies Co., Ltd.*

Context & Objective

- Studies using higher-resolution HMDs like HTC Vive Pro currently not available → important to investigate influence of increased screen resolution on 360° video QoE (→ future HMD devices will be of higher resolution)
- Compare influence of a higher-resolution HMD like HTC Vive Pro (2880x1600 pixel in total) on 360° video QoE to a lower-resolution HMD like HTC Vive (2160x1200 pixel in total)
- Study effect of better integral video quality of the HTC Vive Pro on the discrimination power of subjective ratings
- Evaluate difference in perceived quality for entertainment-type of 360° content in 4K/6K/8K resolution at typical high-quality bitrates → Is 8K resolution really providing a considerably better quality than 4K/6K?
- Get to know, which areas of the video people are focusing on while watching them → Important issue for producers of 360° contents
- Are there differences in head rotation behavior between a) the different HMDs and b) the single quality levels

https://www.vive.com/filer/sharing/1529456235/9112/

https://images-na.ssl-images-amazon.com/images/I/81Q1I3RtZQL._SX385_.jpg

Publications

- **A. Singla** and S. Fremerey and W. Robitza and A. Raake, “Measuring and comparing QoE and simulator sickness of omnidirectional videos in different head mounted displays”, QoMEX, May 2017
- **A. Singla**, A. Raake, W. Robitza, P. List, B. Feiten, “AhG8: Subjective Quality Evaluation for Omnidirectional (360°) Videos”, JVET-G0152, 7th Meeting, July 2017
- **A. Singla** and S. Fremerey and W. Robitza and P. Lebreton and A. Raake, “Comparison of Subjective Quality Evaluation for HEVC Encoded Omnidirectional Videos at Different Bit-rates for UHD and FHD Resolution”, Thematic Workshops of ACM MM, October 2017
- **A. Singla**, S. Fremerey, A. Raake, P. List, B. Feiten, “AhG8: Measurement of User Exploration Behavior for Omnidirectional Videos with a Head Mounted Display”, JVET-H0050, 8th Meeting, October 2017
- **A. Singla** and W. Robitza and A. Raake, “Comparison of subjective quality evaluation methods for omnidirectional videos with DSIS and Modified ACR”, HVEI, January 2018
- S. Fremerey, **A. Singla**, K. Meseberg, A. Raake, “AVTrack360: An open Dataset and Software recording people’s Head Rotations watching 360° Videos on an HMD” ACM Multimedia Systems Conference (MMSys), June 2018