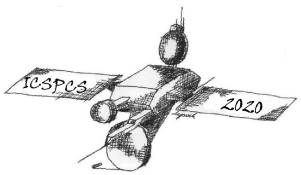
Comparison of ACR Methods for 360° Video Quality Assessment Subject to Participants' Experience with Immersive Media

Majed Elwardy Yan Hu Hans-Jürgen Zepernick
Thi My Chinh Chu Veronica Sundstedt

Blekinge Institute of Technology, Sweden

14^{th} International Conference on Signal Processing and Communuication Systems (ICSPCS)
Adelaide, South Australia, 14.12.2020 (Online)



Introduction (1)

- Immersive media such as virtual reality (VR) content shown on head-mounted displays (HMDs) have become more popular among consumers.
- 360° videos on HMDs and their subjective and objective quality assessment have seen an increased interest in recent years.
- Subjective quality assessment of 360° videos on HMDs is needed to obtain a ground truth on the quality as perceived by users.

Introduction (2)

- Due to the lack of standardized methods, the absolute category rating (ACR) with hidden reference (ACR-HR) method for conventional videos has been used for subjective quality assessment of 360° videos on HMDs [ITU-T P.910, 2008].
- A modified ACR with hidden reference (M-ACR-HR) method tailored for assessing 360° videos on HMDs has recently been proposed [Singla et al., 2017].
- In this paper, we compare the ACR-HR and M-ACR-HR method regarding subjective quality assessment of 360° videos on HMDs.

Contribution

- Subjective experiments were conducted for 360° video quality assessment shown on an HTC Vive Pro HMD using the ACR-HR and M-ACR-HR method.
- Three classes of experience levels with immersive media on HMDs are considered in the comparison of the experimental data: Experts, sometimes used, never used.
- A detailed statistical analysis is provided supporting the comparison in terms of average rating times, MOS and standard deviation (STD), t-tests between opinion scores of both methods, and answers to a simulation sickness questionnaire (SSQ).

360° Reference and Test Videos (1)

- Four 360° video scenes were selected from the VQA-ODV database in 7680×3840p (8K) resolution [VQA, 2019].
- The 360° reference videos of 10 s duration were produced.
- By down-sampling the 8K reference videos with the bi-cubic scaling algorithm, additional reference videos with 6K, 4K, 2K and optimal resolution (OR) of 3600×1800p recommended in [Zhang et al., 2018] for the HTC Vive Pro HMD were generated.

360° Reference and Test Videos (2)

- The reference videos is encoded into HEVC/H.265 format with five different quantization parameters (QPs), i.e., QP = 22, 27, 32, 37, and 42.
- A total of 120 360° videos were generated including the reference videos with different resolutions and the corresponding test videos with different QPs.

360° Reference and Test Videos (3)

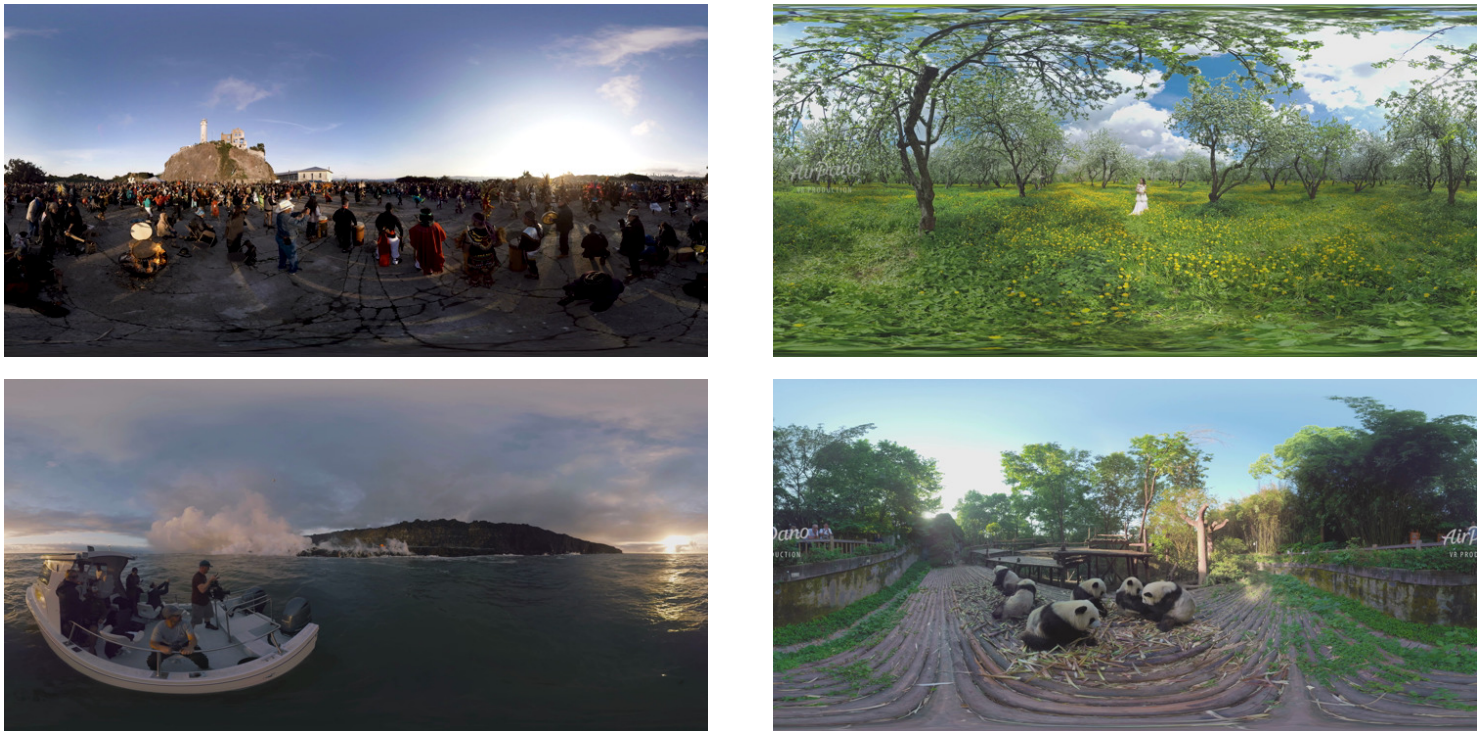


Figure 1: Sample frames of the four 360° video scenes in equirectangular projection from the sphere to the plane.

Software Suite and Technical Equipment

- The test platform was developed using the Unity 3D game engine.
- The test platform includes graphical user interfaces for giving the opinion scores to the stimuli by using the HTC Vive controller.
- The Unity build-in random function was used to randomize the video presentation.

Procedures

- An introduction to the tasks of the experiment, safety instructions, and risks with using HMDs were provided to the participants.
- A visual acuity and color vision test were conducted.
- The participants that passed the vision tests and signed the consent form commenced with a training session.
- All test sessions were concluded with answering a simulator sickness questionnaire (SSQ) and general comments of the participants to the experiments were collected.

Test Methods

- ACR-HR: Each hidden reference and test video is shown once and its quality is then directly rated using a five-level quality rating scale: (5) Excellent, (4) Good, (3) Fair, (2) Poor, (1) Bad as recommended in [ITU-T P.910, 2008, ITU-T P.915, 2016].
- M-ACR-HR: Each hidden reference and test video is shown twice separated by a 3 s mid-grey screen between the two presentations and its quality is then directly rated the same as in the ACR-HR method.

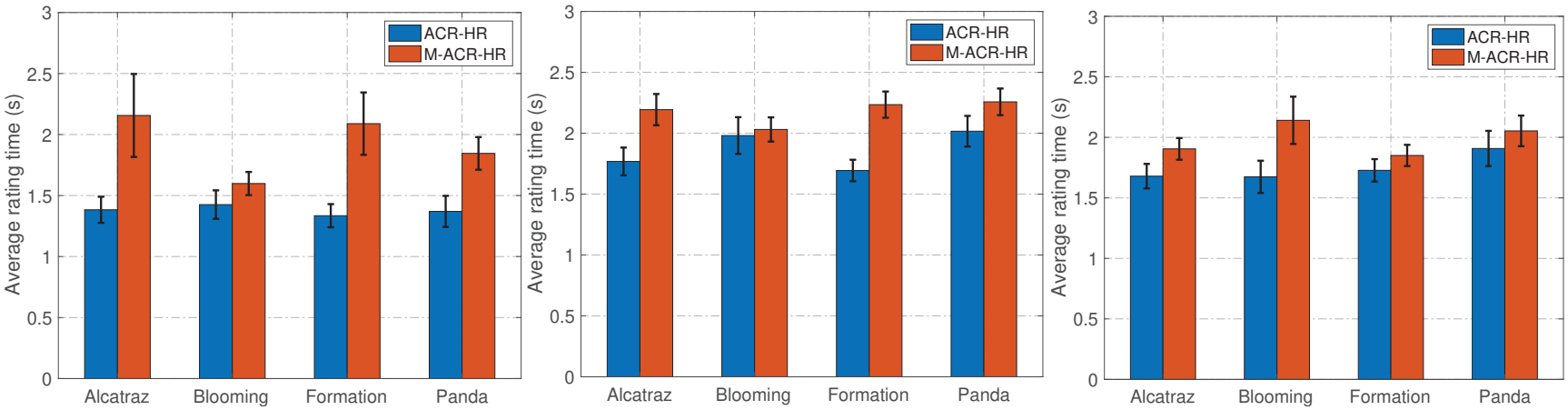
Participants

- ACR-HR method: A total of 32 volunteers were recruited with two of them failing the vision tests.
- M-ACR-HR method: A total of 33 volunteers were recruited with three of them failing the vision tests.
- A pilot study was also conducted for both methods engaging 5 experts.

Table 1: Screening Results for Consistency of Opinion Scores

Method	Number of Valid Participants		
	Sometimes used	Never used	Total
ACR-HR	15	11	26
M-ACR-HR	13	13	26

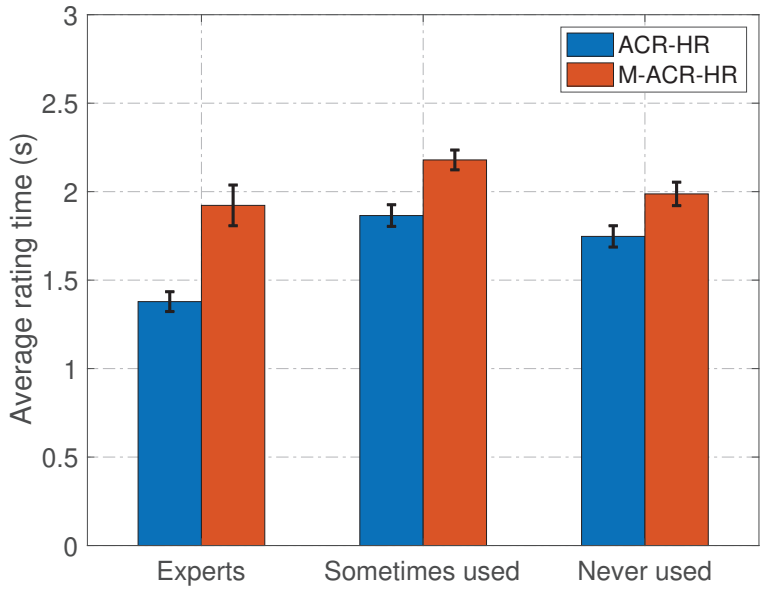
Average Rating Time (1)



(a) Experts (b) Somtimes used (c) Never used

Figure 2: Average rating times and 95% CIs for the ACR-HR and M-ACR-HR method subject to participants' experience with immersive media on HMDs.

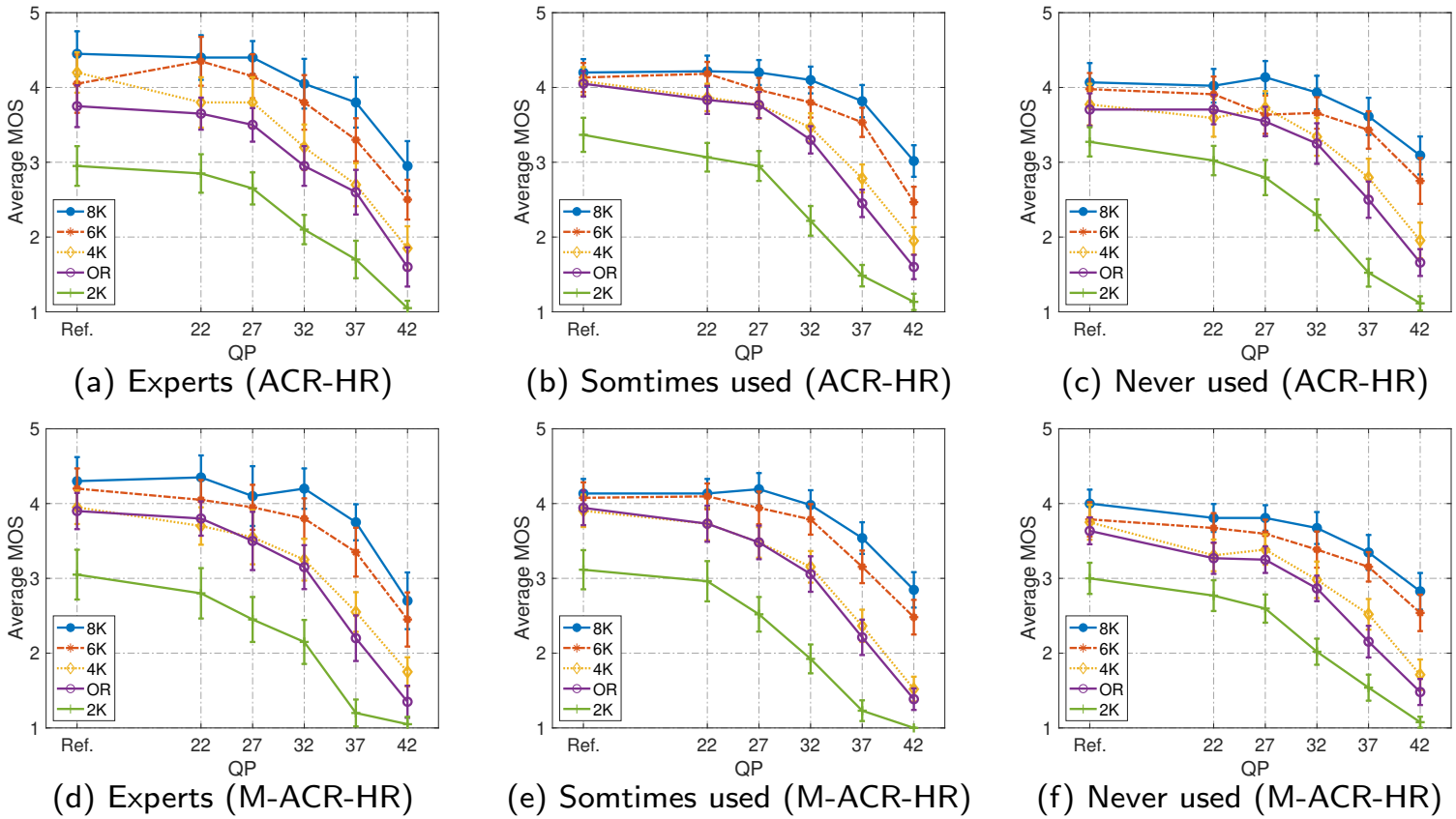
Average Rating Time (2)



Average rating time over all four scenes

Figure 3: Average rating times and 95% CIs for the ACR-HR and M-ACR-HR method over all four scenes.

MOS and Standard Deviation (1)

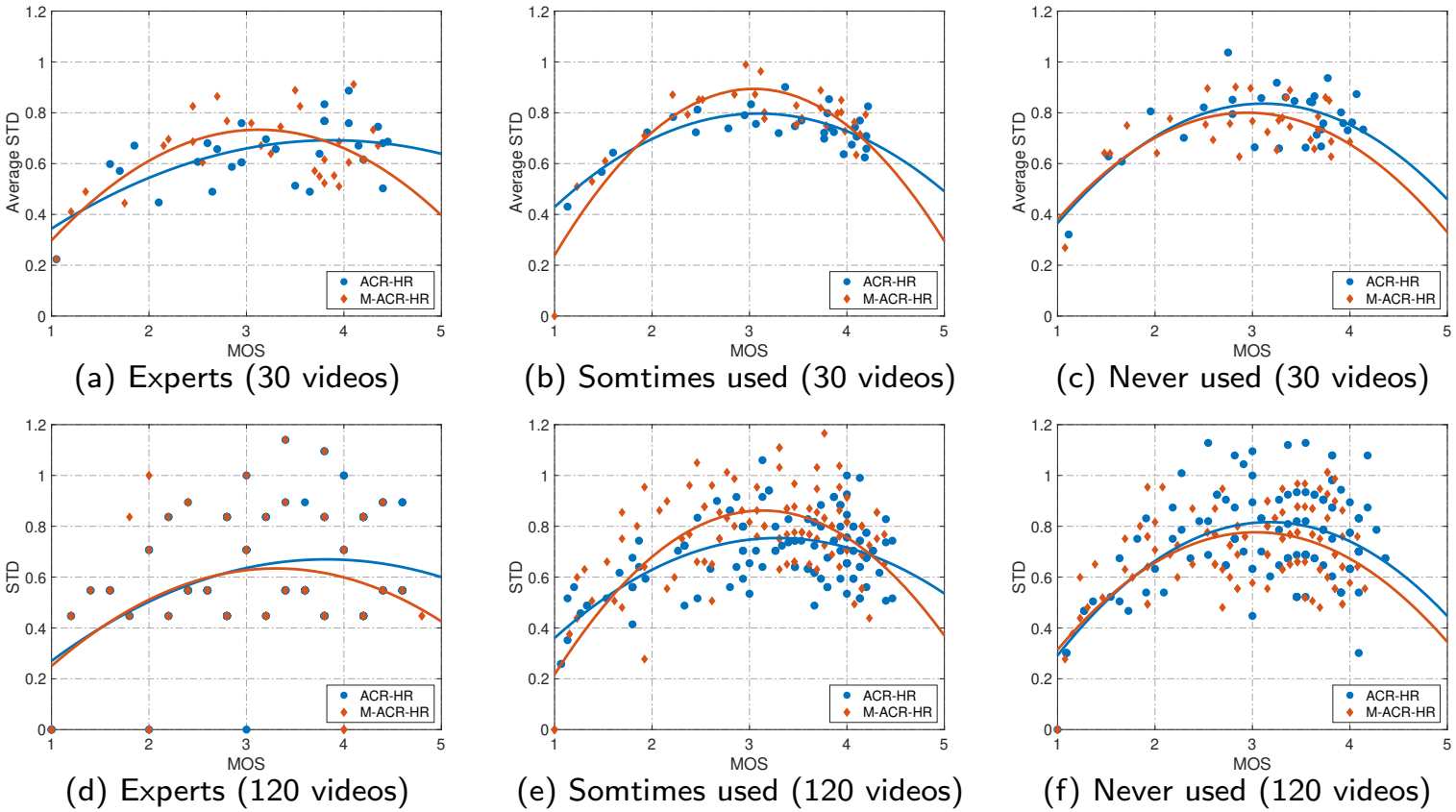


(a) Experts (ACR-HR) (b) Somtimes used (ACR-HR) (c) Never used (ACR-HR)

(d) Experts (M-ACR-HR) (e) Somtimes used (M-ACR-HR) (f) Never used (M-ACR-HR)

Figure 4: Average MOS and 95% CIs for the ACR-HR and M-ACR-HR method subject to participants' experience.

MOS and Standard Deviation (2)

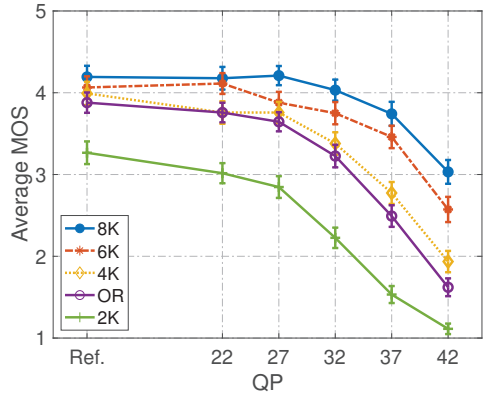


(a) Experts (30 videos) (b) Somtimes used (30 videos) (c) Never used (30 videos)

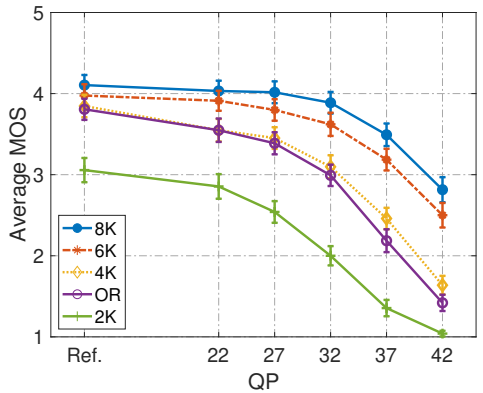
(d) Experts (120 videos) (e) Somtimes used (120 videos) (f) Never used (120 videos)

Figure 5: Average standard deviation and polynomial fit over the four different scenes (30 videos per scene) and STD for 120 videos.

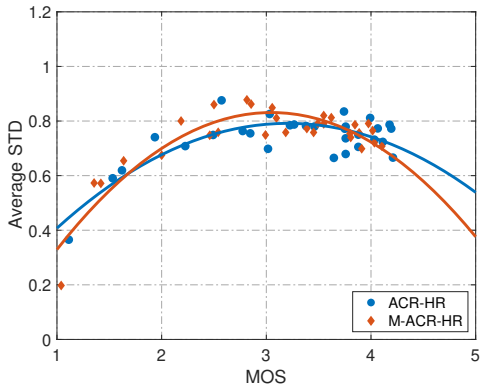
MOS and Standard Deviation (3)



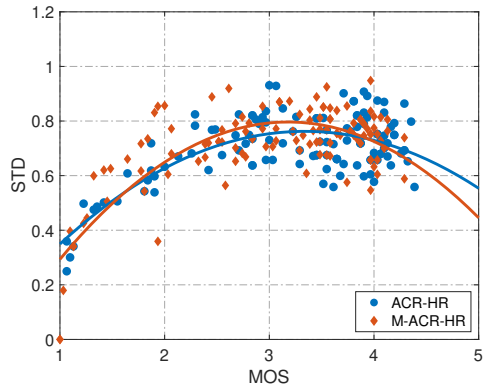
(a) ACR-HR



(b) M-ACR-HR



(c) Average standard deviation over the four scenes



(d) Standard deviation for 120 videos

Figure 6: Overall statistics without differentiation among participants' experience.

t-tests

- The Bonferroni-corrected t-test was performed between each resolution-QP pair for each of the two methods.
- Given the 120 different 360° videos, there exist $120 \times 119/2 = 7140$ possible combinations of resolution-QP pairs.

Table 2: Bonferroni-corrected t-test for 7140 Possible Resolution-QP Pairs and Significance Level $\alpha = 0.05$

	Overall	Experts	Sometimes used	Never used
ACR-HR	3427	1183	2536	1513
M-ACR-HR	3512	1850	2444	1967

Simulator Sickness Questionnaire (SSQ)

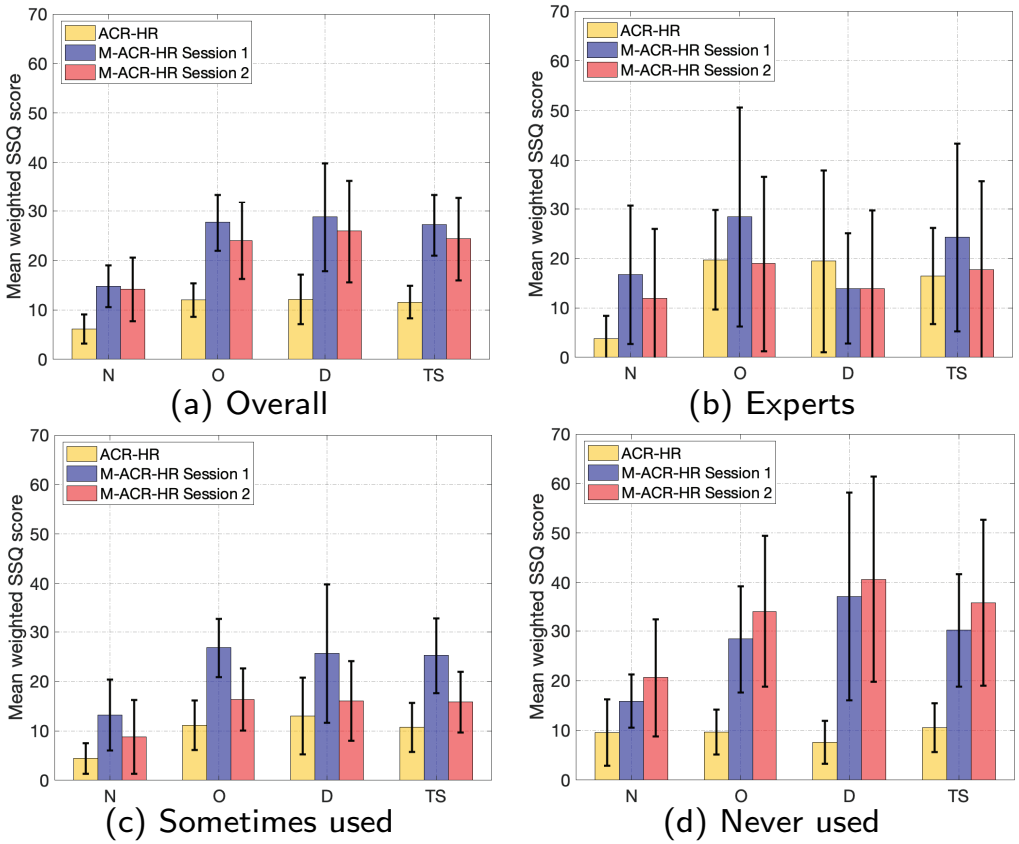


Figure 7: Statistical analysis of symptom cluster scores in terms of mean weighted SSQ scores and their 95% CI.

Conclusions

- Average rating times to cast an opinion score are lower for the ACR-HR method irrespective of the scene shown and participants' experience.
- The progression of MOS values for different resolutions versus QPs is similar for both methods irrespective of participants' experience.
- A Bonferroni-corrected t-test revealed that the M-ACR-HR method is, in general, more reliable compared to the ACR-HR method.
- Analysis of the SSQs shows that discomfort is kept significantly lower for the ACR-HR compared to the M-ACR-HR method.

Acknowledgments

- This work was supported in part by The Knowledge Foundation, Sweden, through the ViaTecH project (Contract 20170056).
- The study has been granted ethical approval by the Regional Ethics Review Board in Lund, Sweden (Dnr. 2018/863).
- We thank all volunteers who generously shared their time to participate in the subjective experiments. Special thanks go to Francisco Lopez Luro and Diego Navarro for their advice given to the development of the test platform.

References

(2017 (accessed Apr. 27, 2019)). *VQA-ODV*. Beihang University, School of Electronic and Information Engineering, Beijing, China. https://github.com/Archer-Tatsu/VQA-ODV.

ITU-T P.910 (2008). *Subjective Video Quality Assessment Methods for Multimedia Applications*. Int. Telecommun. Union.

ITU-T P.915 (2016). *Subjective Assessment Methods for 3D Video Quality*. Int. Telecommun. Union.

Singla, A., Fremerey, S., Robitza, W., Lebreton, P., and Raake, A. (2017). Comparison of Subjective Quality Evaluation for HEVC Encoded Omnidirectional Videos at Different Bit-rates for UHD and FHD Resolution. In *Proc. Thematic Workshops of ACM Multimedia*, Thematic Workshops '17, pages 511–519, Mountain View, CA, USA.

Zhang, Y. et al. (2018). Subjective Panoramic Video Quality Assessment Database for Coding Applications. *IEEE Trans. on Broadcast.*, 64(2):461–473.

Thank you !

Majed Elwardy
majed.elwardy@bth.se

(Please contact me if you have questions)