VQEG

IMG Test Phase 1 - Short Sequences Results and Outcomes

VQEG Plenary Meeting, December 2020

Femi Adeyemi-Ejeye, Federica Battisti, Kjell Brunnström, Marco Carli, Pablo César, Zhenzhong Chen, Natalia Cieplińska, Carlos Cortés, César Díaz, Stephan Fremerey, Narciso García, Jesús Gutiérrez, Omar Hamsis, John Hedlund, Frank Hofmeyer, Yaosi Hu, Lucjan Janowski, Dawid Juszka, Peter Lambert, Mikołaj Leszczuk, Pramit Mazumdar, Marta Orduna, Pablo Pérez, Alexander Raake, Ashutosh Singla, Glenn Van Wallendael, Irene Viola

IMG tests: Quality Assessment of 360-degree Video

- Global target of the test plan: https://www.its.bldrdoc.gov/vqeg/projects/immersive-media-group.aspx
  - Design and execute a **cross-lab test** where we can assess and validate **subjective evaluation methodologies** for 360-VR video.
  - Contribute to the **standardization** of such methodologies under ITU-T Q13/12.
  - Generate a **dataset** of subjectively assessed content for future research.
- What Phase 1 covers (completed):
  - Assessment of **short sequences** (<=30s), in the spirit of ITU-R BT.500, ITU-T P.910, etc.
  - Assessment of **video quality** and **simulator sickness**.
- What is not covered in Phase 1 (will be in Phase 2)
  - Assessment of long sequences (several minutes).
  - Assessment of presence.

IMG tests: Quality Assessment of 360-degree Video

- The outcomes from the tests of Phase 1 has resulted in:
  - Two contributions to ITU SG12/Q13 (meetings in Apr. and Sep. 2020).
  - The approval in mid-October of the new **ITU-T Recommendation P.919** (ex P.360-VR): "Subjective test methodologies for 360-degree video on head-mounted displays".
  - Currently available in 'pre-published' form at https://www.itu.int/rec/T-REC-P.919/en (for TIES account holders). Published version to be available soon.
- This presentation shows some of the results that support the recommendation.

IMG tests: Quality Assessment of 360-degree Video

Objectives and test conditions

- **Audiovisual Quality**
  - Test methodology
    - ACR vs DCR (Degradation Category Rating - Double Stimulus).
  - Sequence duration
    - 10s vs 20s.
    - 20s vs 30s.
    - 10s vs 30s.
  - Test setup
    - HTC Vive vs Samsung Gear VR vs HTC Vive Pro.
    - Scoring app vs “Saying out lout”.
    - Tethered vs Untethered.
- **Simulator sickness**
  - When/how to assess simulator sickness.
  - Short vs. long questionnaires.

Test Setup

Distribution of test conditions, participant labs and observers

ID	Test condition	Lab	HMD1	HMD2	Comment	Observers (female/ male)
A	ACR: 10s vs 20s	Wuhan	Vive			30 (15/15)
B	ACR: 20s vs 30s	AGH	O. Rift			39 (13/26)
C	DCR: 10s vs 20s	Roma3	Vive			30 (8/22)
D	DCR: 20s vs 30s	CWI	O. Rift			28 (14/14)
E	HMD vs HMD (ACR 20s)	Nokia	GearVR	Vive Pro	Mobile vs desktop	60 (25/35) 40 per HMD
F	HMD vs HMD (ACR 20s)	UPM	Vive	Vive Pro	Low res vs. High res	
G	HMD vs HMD (ACR 20s)	Ghent	Vive Pro	Vive Pro	Tethered vs Untethered	30 (4/26)
H	With vs without audio (ACR 20s)	RISE	Vive			28 (16/12)
I	Scoring interface vs voice (ACR 20s)	TUI	Vive Pro			29 (14/15)
J	ACR: 10s vs 30s	Surrey	Vive			31 (10/21)

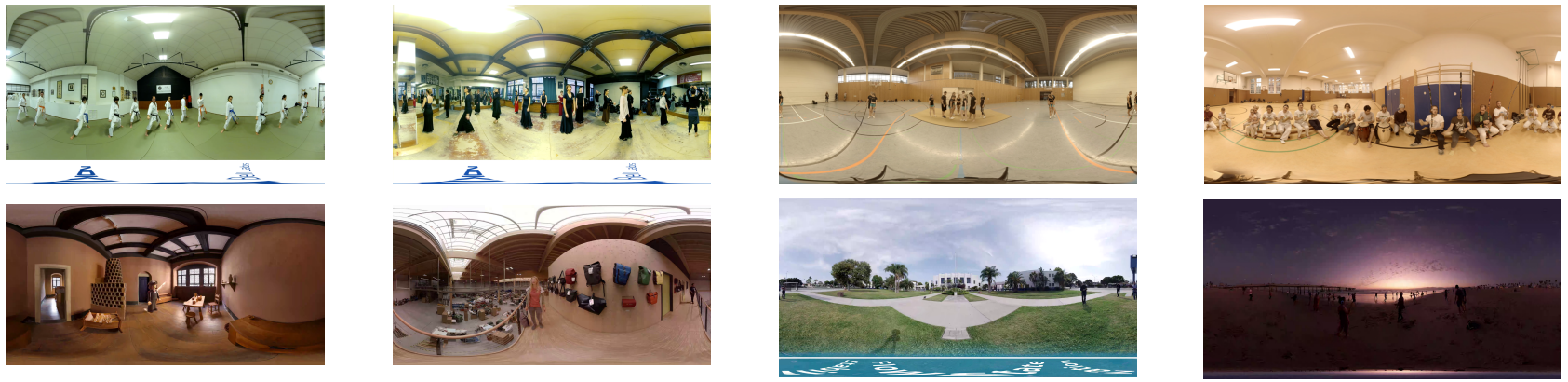
Minimum number of observers = 28. Supported by statistical analyses based on statistical power (by Irene Viola) and on the VQEGNumSubjTool* (by Kjell Brunnström).

*W. Robitza and K. Brunnström, Nov 2019. [Online]. Available:https://slhck.shinyapps.io/number-of-subjects

Test Setup

SRCs

- 8 Raw sources:
  - 4K (some available in 8K), equirectangular, monoscopic.
  - 10, 20 and 30 second cuts.
  - Dojo Zentrum, Flamenco*, Cheer Leading, Brazil Music*, Vaude*, Luther*, OculusMotion*, OculusBeach.
  - From Nokia, TU Ilmenau, VSense, Oculus.
  - * Reduced sub-set for long test session (DCR, 30 seconds sequences...).



Test Setup

HRCs

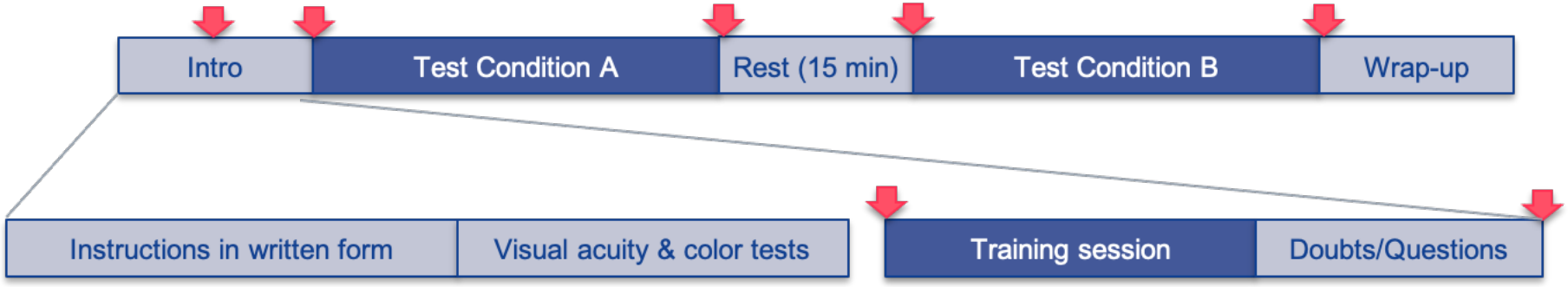
- 8 coding configurations (including Hidden Reference), HEVC, fixed QP:
  - 4 homogeneous QPs: 15 (HR), 22, 32, 42.
  - 4 non-homogeneous QPs (tiles).
  - Reduced test set for time-limiting conditions (DCR, 20s, 30s, etc.): **Removed for 6 HRC test.

#Tiles	Transition	ROI	QPs							
8x5**	Smooth	90^{O}	42	37	32	22	22	32	37	42
6x3	Smooth	120^{O}	42	32	22	22	32	42		
8x5**	Abrupt	180^{O}	42	42	22	22	22	22	42	42
6x3	Abrupt	120^{O}	37	37	22	22	37	37		

Test Setup

Test session

- Each subject tests *the same PVSs* under two different conditions.
  - Subjects sitting in a swivel chair.
- Each condition is tested in one *active period.*
  - After each active session, we should have a rest period of 15 min.
- Red arrows: score simulator sickness.

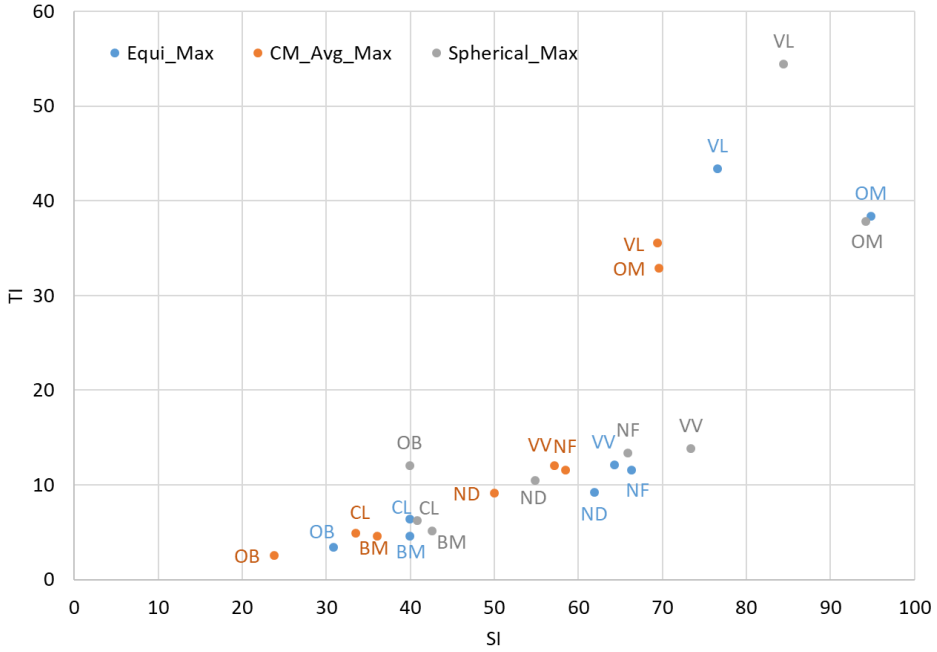


- Each subject's participation is limited to 1.5 hours, of which no more than 50 minutes is spent rating stimuli, and no more than 25 minutes continuously.

Test Setup

Content characterization: SI/TI

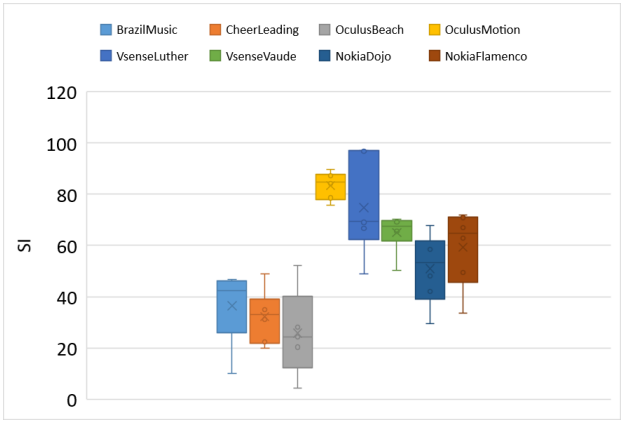
- Computed according to ITU-T P.910, with code from VQEG Tools on equirectangular (ER), cubemap (CM) [1], and spherical projections [2] → To account for projection distortions.

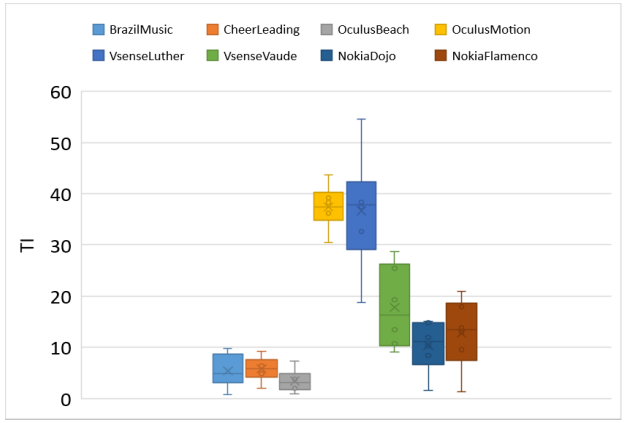


High correlation between ER, CM and Spherical:

Pearson			
SI (Max)	**Equi**	**CM**	**Spherical**
Equi	1	0.9638	0.9638
CM	0.9638	1	0.9543
Spherical	0.9638	0.9543	1
TI (Max)	**Equi**	**CM**	**Spherical**
Equi	1	0.9968	0.8857
CM	0.9968	1	0.9174
Spherical	0.8857	0.9174	1

Variability on the complexity of the CM faces for some contents:
Single complexity value may not be very informative [1].





[1] *F. De Simone et al.*, "Complexity measurement and characterization of 360-degree content", *HVEI2019*.
[2] *Wang et al.*, "Equirectangular Projection Oriented Intra Prediction for 360-Degree Video Coding", *VCIP2020*

Test Results

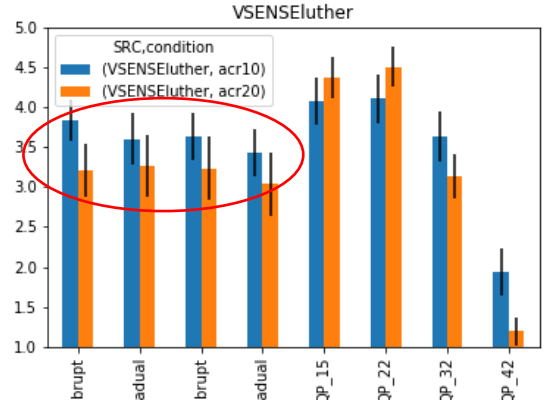
Duration of stimuli & test methodologies

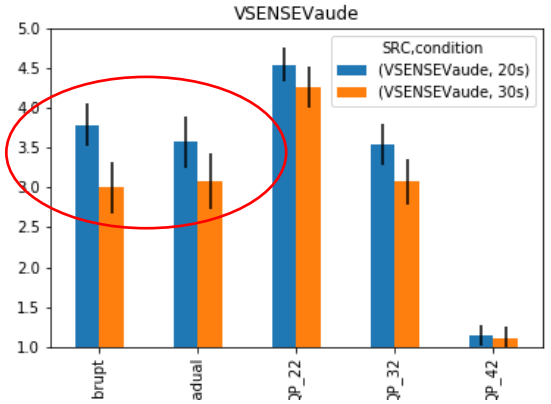
- Test sequences of 10, 20 and 30 seconds were considered.

- Generally, **very similar results for audiovisual quality** were obtained in the three cases.
  - Differences observed in videos with new scenes appearing after 10/20 seconds, or where the video properties changed.
  - Differences mainly observed with non-uniform coding schemes.
  - Further analyses ongoing.

- Similar results observed for ACR and DCR methodologies.

- Ongoing analyses on subject bias and outlier detection methods…





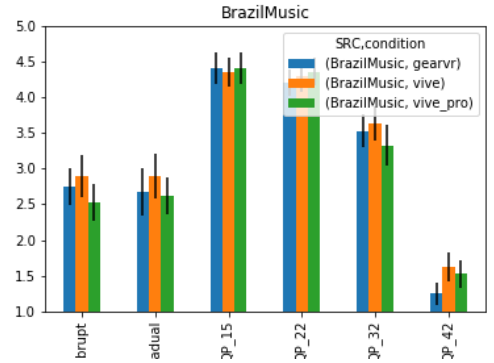
Test Results

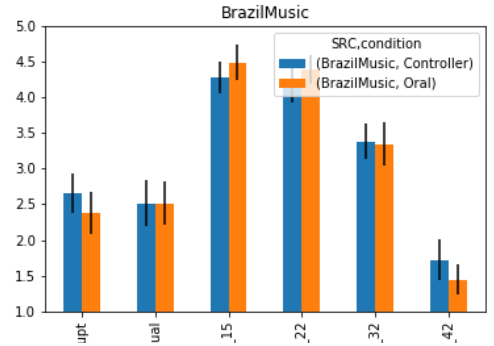
Equipment and audio considerations

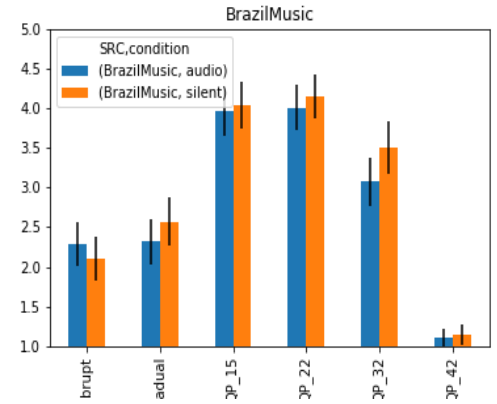
- Very similar results found in audiovisual quality ratings with Samsung Gear VR, HTC Vive, HTC Vive Pro (tethered and untethered).
  - Ongoing analyses on exploration behavior differences.

- Very similar results found in audiovisual quality ratings scoring with the VR app (Miro360*) and with verbal voting.

- Very similar results found in audiovisual quality ratings with sequences with and without audio.





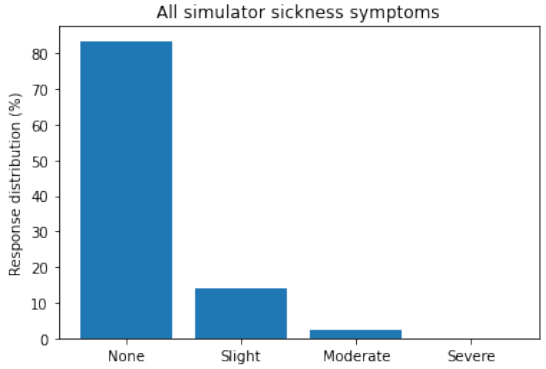


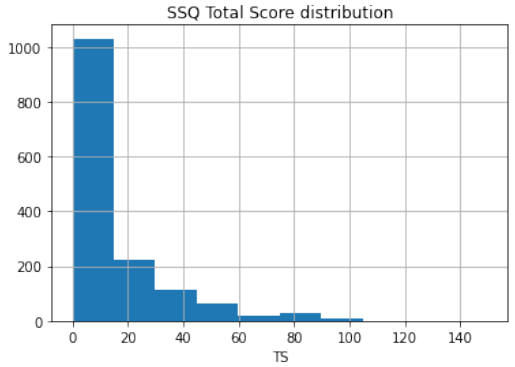
*C. Cortes, P. Perez, and N. Garcia, "Unity3D-based app for 360VR subjective quality assessment with customizable questionnaires," *IEEE International Conference on Consumer Electronics*, Sep. 2019.

Test Results

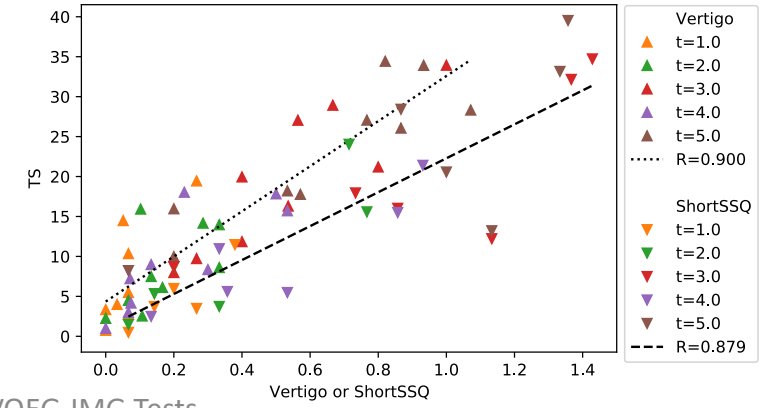
Evaluation of Simulator Sickness

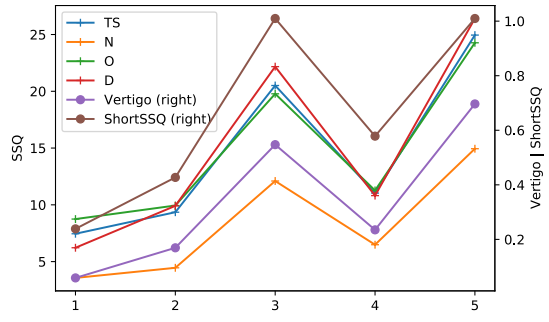
- Evaluated 5 times during the test session:
  - SSQ [Kennedy 1993] (long, 16 symptoms), Vertigo Scale [Perez 2018] (single question), and Short SSQ (single question) [Tran 2017].
  - In general, low simulator sickness (some slight/moderate symptoms).
  - Evidenced effect of the break.
- SSQ results correlate with single-question results:
  - If we aggregate results from all subjects, they contain enough information.
- VRSQ [Kim et al. 2018] can be used as a sub-sampling of the long SSQ: 9 symptoms, 2 factors.





	Question	Dv	VRSQ Ov	TSv
SSQ	D	0.910692	0.729877	0.872603
	O	0.787438	0.959722	0.971799
	TS	0.853195	0.895551	0.958025





By Pablo Pérez.

IMG tests: Quality Assessment of 360-degree Video

- The outcomes form the tests of Phase 1 has resulted in:
  - Two contributions to ITU SG12/Q13.
  - The approval in mid-October of the new **ITU-T Recommendation P.919** (ex P.360-VR): “Subjective test methodologies for 360-degree video on head-mounted displays”.
  - Currently available in ‘pre-published’ form at https://www.itu.int/rec/T-REC-P.919/en (for TIES account holders). Published version to be available soon.
  - One contribution (M54398) to MPEG AhG on Quality of Immersive Media at the MPEG meeting in July 2020.
  - Joint journal paper to be submitted soon, including the publication of the annotated dataset.

VQEG

IMG Test Results
Phase 1 - Short Sequences

VQEG Plenary Meeting, December 2020

Femi Adeyemi-Ejeye, Federica Battisti, Kjell Brunnström, Marco Carli, Pablo César, Zhenzhong Chen, Natalia Cieplińska, Carlos Cortés, César Díaz, Stephan Fremerey, Narciso García, Jesús Gutiérrez, Omar Hamsis, John Hedlund, Frank Hofmeyer, Yaosi Hu, Lucjan Janowski, Dawid Juszka, Peter Lambert, Mikołaj Leszczuk, Pramit Mazumdar, Marta Orduna, Pablo Pérez, Alexander Raake, Ashutosh Singla, Glenn Van Wallendael, Irene Viola