

IMG Work plan Report

Phase 1 - Short Sequences

Jesús Gutiérrez

Grupo de Tratamiento de Imágenes (GTI)

Universidad Politécnica de Madrid (UPM)

jesus.gutierrez@upm.es

Pablo Pérez

Nokia Bell-Labs

Madrid, Spain

Context

- **Mission:** Quality assessment of immersive media, including virtual reality, augmented reality, plenoptic, stereoscopic 3DTV, multiview...
- **Goals:** Baseline quality assessment of immersive systems
 - **Datasets** of immersive media content
 - **Subjective test methods**, QoE **guidelines**, presentation requirements, etc.
- **Technologies:**
 - 360-degree content / Virtual Reality
 - Augmented/Mixed reality
 - 3D, FVV, multiview technologies, including full parallax
 - Light field processing also called plenoptic
- Email reflector: img@vqeg.org

Context

- Discussions, contributions, and presentations in the f2f meetings are more than welcome covering any immersive media technology.
- In March 2018 (meeting in Madrid) it was decided to:
 - Push joint work.
 - Focus on some more specific uses cases for the joint work
 - Monthly audio calls
- Identified uses cases of interest for the participant labs:
https://docs.google.com/spreadsheets/d/1FgnRXjiU4td_KwkeKwGvy652nCDBvrAvkSug6HZivZ4/edit#gid=0
- Decided in July 2018 to consider the use case of **subjective quality evaluation of 360-degree video** for joint work.

Quality assessment for 360° content

- Working on a **test plan**:
https://docs.google.com/document/d/18L_Q7vyzEsoAzbRsTKBTtMrwfjISMf0HUGofPqOOJl/edit?usp=sharing
- Specific proposals for the test plan on evaluating quality for:
 - **Short sequences**: Length of sequences, ACR vs DCR, Influence of HW, typical artifact (coding, projections, stitching, etc.)...
 - **Long sequences**: immersiveness, presence, simulator sickness, etc.
- Collaboration **VQEG-IMG and ITU-T SG12 Q13**:
 - Recommendation **P.360-VR**: *Subjective test methodologies for 360 degree video on HMD*
 - Collaboration aligned with the joint work on the VQEG-IMG test plan.

In the last meeting in Shenzhen (Oct. 2019) it was decided to kick-off Phase 1 (short sequences)

Phase 1. Short Sequences

Contribution to P.360-VR

- Global target of the test plan
 - Design and execute a **cross-lab test** where we can assess and validate **subjective evaluation methodology** for 360-VR video
 - Contribute to the **standardization** of such methodology under ITU-T P.360-VR (Q13/12)
 - Generate a **dataset** of subjectively assessed content for future research
- What Phase 1 covers
 - Assessment of **short** sequences (≤ 30 s), 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**.

Phase 1. Short Sequences

Main objectives - Contribution to P.360-VR

Methodology:

1. Recommend:
 - a. Implementation of ACR and DCR methodologies for 360 videos
 - b. Sequence duration
2. Study of influence factors:
 - a. HMD
 - b. Source content characteristics
 - c. Uniform vs non-uniform artifacts

SSQ:

1. Recommend:
 - a. One multi-item questionnaire (SSQ or derivation from it), and/or one single-question item
2. Write some guidelines about:
 - a. When/how to assess simulator sickness
 - b. When can you use a single-question vs. full questionnaire
 - c. How to process the results / what to do with them

Test Setup



Research Questions - Test conditions

- **Video Quality**

- Test methodology
 - ACR vs DCR (Degradacion 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 loud”
 - Tethered vs Untethered

- **Simulator sickness**

- When/how to assess simulator sickness
- Short vs long questionnaire
- Short-SSQ (Tran 2017) vs Vertigo questionnaire (Pérez 2018)

Test Setup



Distribution of test conditions and participant labs

ID	Test condition	Lab	HMD1	HMD2	Comment	Status
A	ACR: 10s vs 20s	Wuhan	Vive			Finished / 31 observers
B	ACR: 20s vs 30s	AGH	O. Rift			Ongoing
C	DCR: 10s vs 20s	Roma3	Vive			Finished / 30 observers
D	DCR: 20s vs 30s	CWI	O. Rift			Finished / 28 observers
E	HMD vs HMD (ACR 20s)	Nokia	GearVR	Vive Pro	Mobile vs desktop	Finished / 60 observers 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	Finished / 31 observers
H	With vs without audio (ACR 20s)	RISE	Vive			Ongoing / 18 observers
I	Scoring interface vs voice (ACR 20s)	TUI	Vive Pro			Finished / 29 observers
J	ACR: 10s vs 30s	Surrey	Vive			Finished / 31 observers

Test Setup

SRCs

- 8 Raw sources
 - 4K (some available in 8K), equirectangular, monoscopic (some available in stereo), static camera.
 - 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

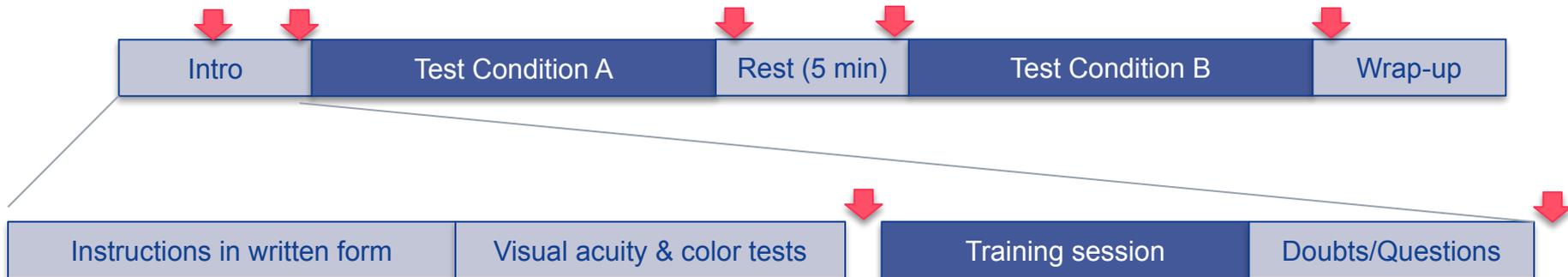
- 8 HRCs (including Hidden Reference), HEVC, fixed QP
 - 4 homogeneous QPs: 15 (HR), 22, 32, 42
 - 4 non-homogeneous QPs (tiles)
 - Reduced test set for time-limiting conditions (DCR, 20s, 30s, etc.): **Removed for 6 HRC test

#Tiles	Transition	ROI	QPs							
8x5**	Smooth	90 ⁰	42	37	32	22	22	32	37	42
6x3	Smooth	120 ⁰	42	32	22	22	32	42		
8x5**	Abrupt	180 ⁰	42	42	22	22	22	22	42	42
6x3	Abrupt	120 ⁰	37	37	22	22	37	37		

Test Setup

Session structure and Measuring methodology

- Each subject tests *the same PVSs* under two different conditions
 - Use within-subject statistics for better statistical power
- 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
- App Miro360: Developed, tested and validated subjective assessment software
 - For desktop (HTC and Oculus) and mobile (GearVR and Daydream)



Test Setup

Session structure and duration

- Total session duration around 1h

Duration	With HMD	What
5 min		Presentation of the test. Instructions must be provided on a written form to subjects.
5 min		Visual screening of the subject.
1 min		Short + SSQ + Pre-test questionnaire
4 min	X	Training session (+ questions)
30 sec		Short + SSQ
~20 min	X	Active period 1
30 sec		Short + SSQ
15 min		Rest period
30 sec		Short + SSQ
~20 min	X	Active period 2
1 min		Short + SSQ + Post-test questionnaire

Test Setup

Simulator Sickness

- In each SS measurement, subjects score both:
 - First, a single question (1-5 scale): always the same (A or B), randomly assigned
 - Second, SSQ [Kennedy 1993]

A. Are you feeling any sickness or discomfort now? Check the appropriate sentence

No problem	No perceptible effect, natural feeling
Light effects	Slight discomfort, but no sickness
Uncomfortable	Moderate discomfort, but tolerable for a while
Unpleasant	Strong discomfort or sickness, but can continue the test
Unbearable	Strong discomfort or sickness, and want to stop test

B. How is your level of dizziness or nausea?

Very dizzy
Dizzy
Slightly dizzy
Not dizzy
Absolutely not dizzy

Test Setup

SSQ: Kennedy 1993

- Circle how much each symptom below is affecting you right now
 1. None
 2. Slight
 3. Moderate
 4. Severe
- General discomfort
- Fatigue
- Headache
- Eye strain
- Difficulty focusing
- Increased salivation
- Sweating
- Nausea
- Difficulty concentrating
- Fullness of head
- Blurred vision
- Dizzy (eyes open)
- Dizzy (eyes closed)
- Vertigo
- Stomach awareness
- Burping

Test Setup

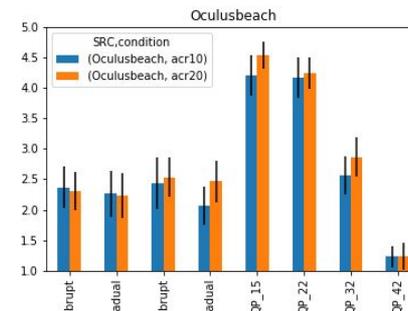
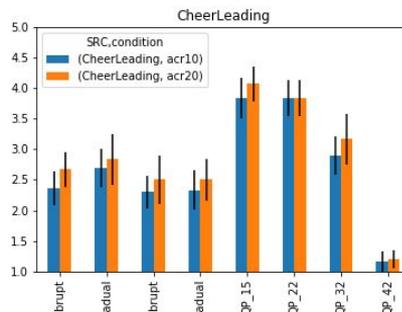
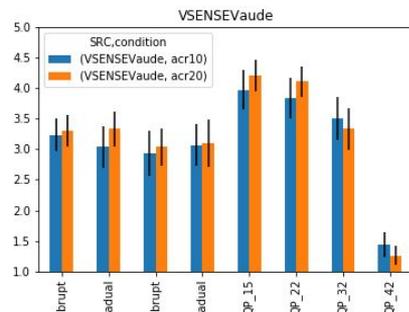
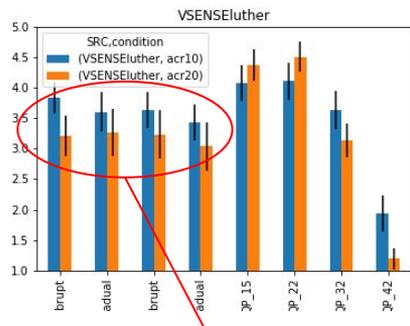
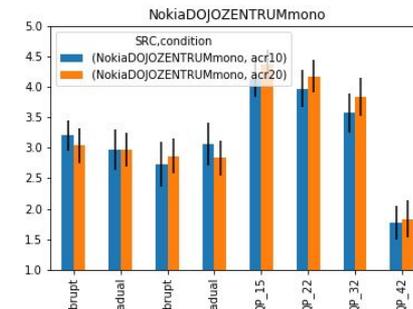
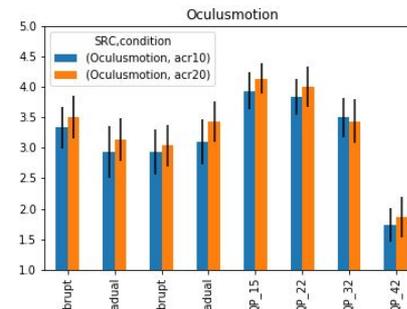
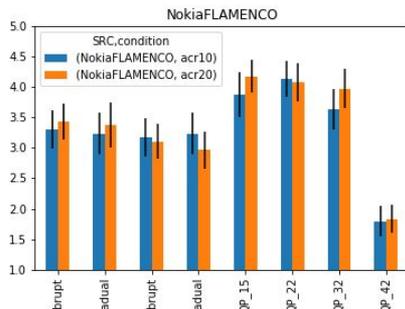
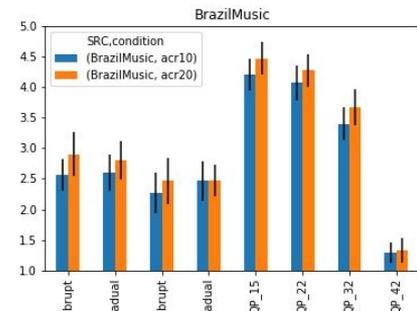
Observers



- At least 28 participants per lab
- Visual screening: Visual acuity and color vision

ID	Test condition	Lab	Total	Female	Male	Age			Experience				
						Min	Max	Avg	First time	Less than 5 times	5 to 20 times	More than 20 times	Every day
A	ACR: 10s vs 20s	Wuhan	30	15	15	20	30	24.5	8	15	7	0	0
B	ACR: 20s vs 30s	AGH											
C	DCR: 10s vs 20s	Roma3	30	8	22	21	57	30.6	7	10	2	8	2
D	DCR: 20s vs 30s	CWI	28	14	14	21	60	27.6	2	12	5	6	3
E	HMD vs HMD (ACR 20s)	Nokia	60	25	35	20	31	23.2	18	32	9	1	
F	HMD vs HMD (ACR 20s)	UPM											
G	HMD vs HMD (ACR 20s)	Ghent	30	4	26	23	45	31.6	3	14	7	5	1
H	With vs without audio (ACR 20s)	RISE											
I	Scoring interface vs voice (ACR 20s)	TUI	29	14	15	20	37	25.9					
J	ACR: 10s vs 30s	Surrey	31	10	21	19	44	25.9	13	12	3	2	1
		Total	238	90	148	19	60	27.04286	51	95	33	22	7
				37.82	62.18				24.40	45.45	15.79	10.53	3.35

Wuhan - ACR: 10s vs 20s - Vive MOS



New scene (different shots) in 20s vs 10s (outdoors, with rain)

Wuhan - ACR: 10s vs 20s - Vive

Main points



- **Preliminary** results from 3-way anova

- Significant effects:

- HRC
- SRC
- **Condition: to verify (SRC Luther)**

- Interactions:

- Condition vs. SRC: Significant
- **Condition vs. HRC: Non significant**
- SRC vs. HRC: Significant

- Further analysis required:

- Check normality → Non-parametric
- Post-hoc tests

- Similar trend in all labs:

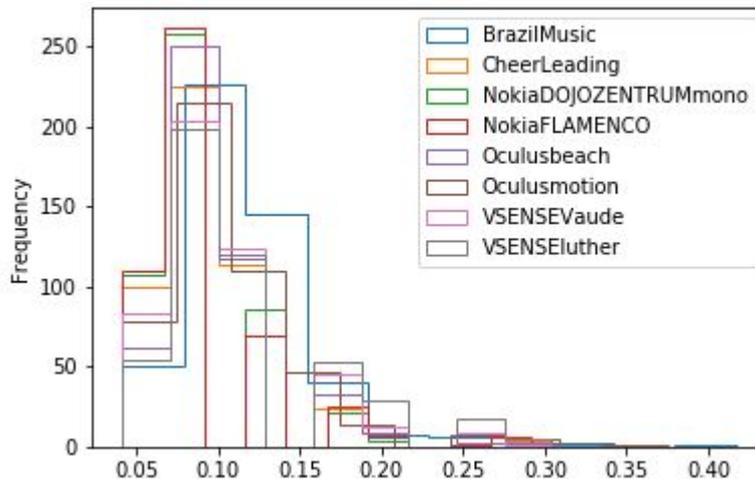
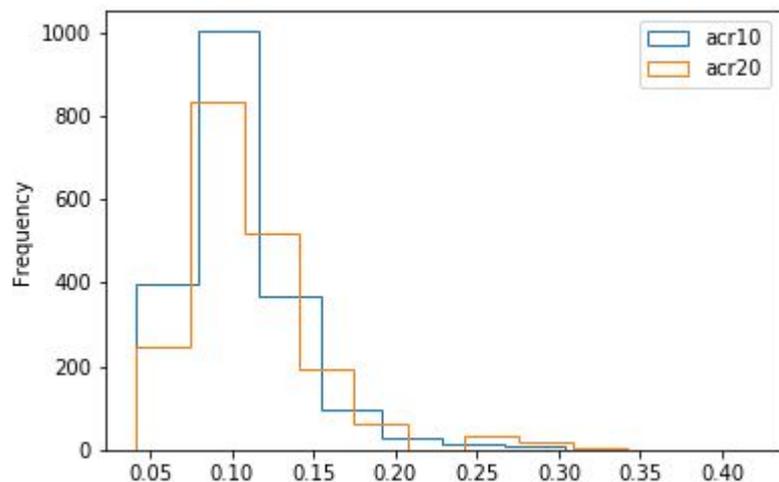
- Acceptable distribution of uniform QPs
- Almost no differences among the 4 non-uniform patterns

```
"          sum_sq      df      F      PR(>F) \
C(condition)      4.887760      1.0      6.540743      1.058253e-02
C(SRC)            211.616406      7.0      40.454651      3.351953e-55
C(HRC)           2319.641406      7.0     443.445222      0.000000e+00
C(condition):C(SRC)      19.539323      7.0      3.735327      4.887297e-04
C(condition):C(HRC)      8.789323      7.0      1.680252      1.090477e-01
C(SRC):C(HRC)        164.781510     49.0      4.500177      9.167298e-23
C(condition):C(SRC):C(HRC)  28.000260     49.0      0.764686      8.842599e-01
Residual          2773.900000    3712.0      NaN          NaN

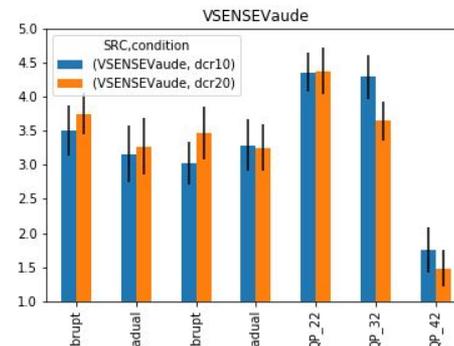
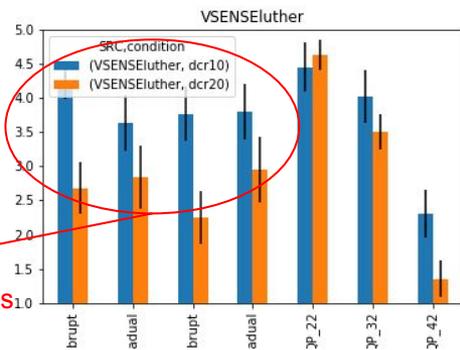
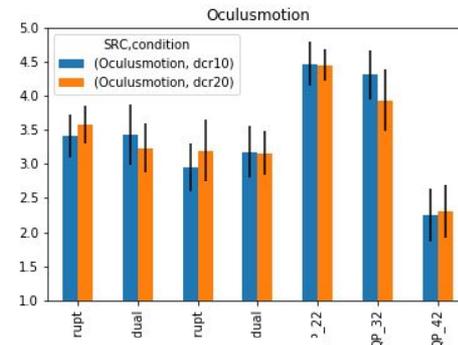
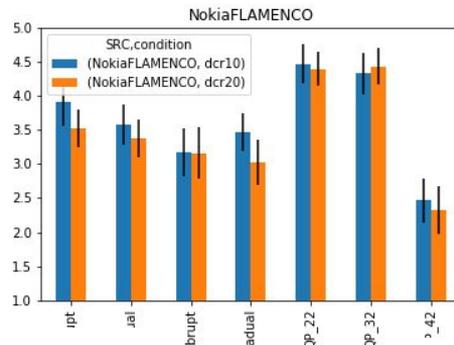
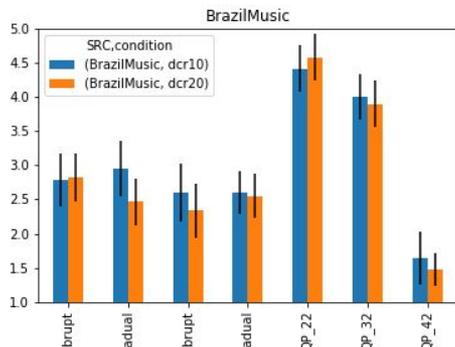
          eta_sq  omega_sq
C(condition)      0.000884  0.000748
C(SRC)            0.038259  0.037308
C(HRC)            0.419377  0.418375
C(condition):C(SRC)  0.003533  0.002587
C(condition):C(HRC)  0.001589  0.000643
C(SRC):C(HRC)     0.029792  0.023168
C(condition):C(SRC):C(HRC)  0.005062 -0.001558
Residual          NaN      NaN "
```

Wuhan - ACR: 10s vs 20s - Vive Exploration

- Histogram of the number (ratio) of visited bins in horizontal (latitude)



Roma3 - DCR: 10s vs 20s - Vive MOS



New scene in 20s vs 10s
(outdoors, with rain)

Roma3 - DCR: 10s vs 20s - Vive

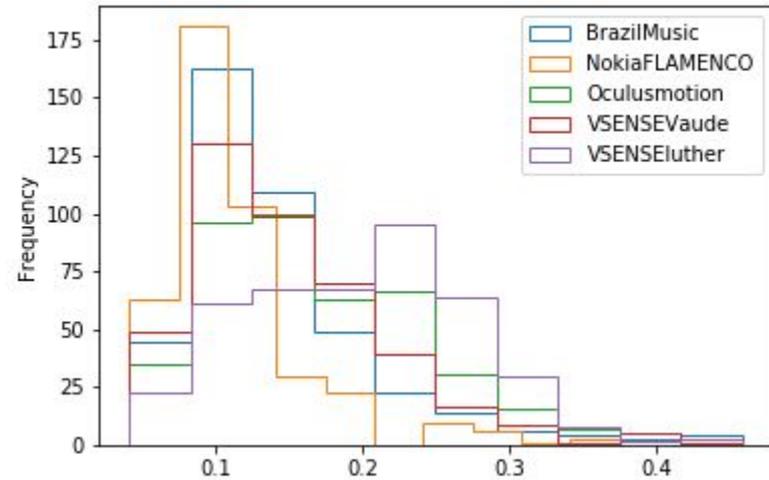
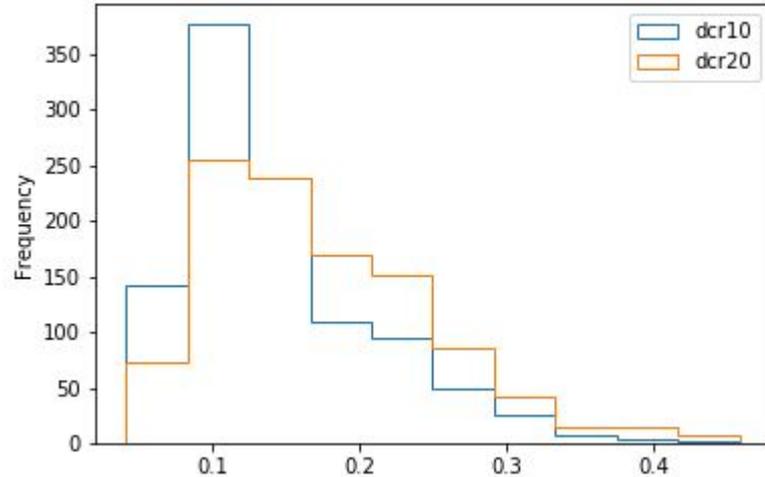
Main points



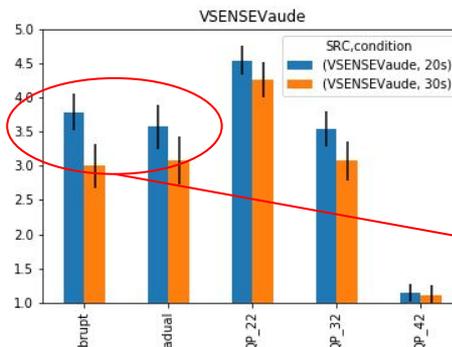
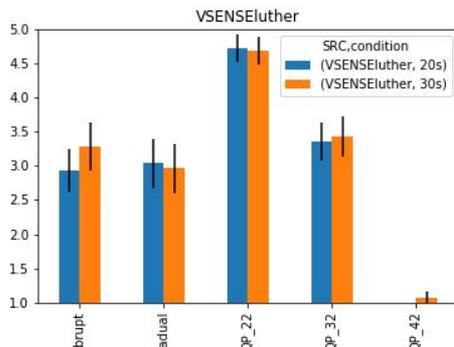
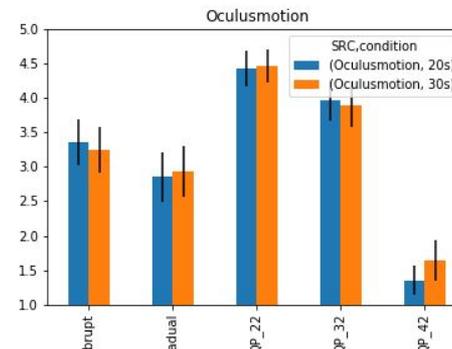
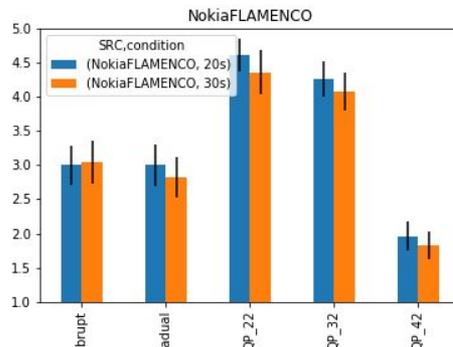
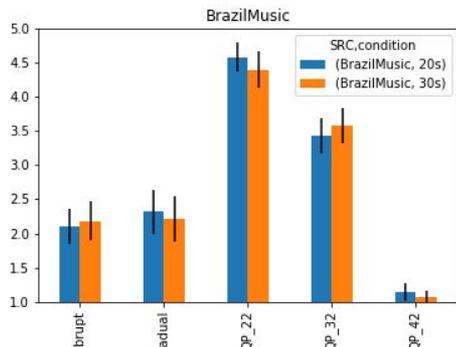
- **Preliminary** results from 3-way anova
 - Significant effects:
 - HRC
 - SRC
 - **Condition: to verify (SRC Luther)**
 - Interactions:
 - Condition vs. SRC: Significant
 - **Condition vs. HRC: Non significant**
 - SRC vs. HRC: Significant
- Further analysis required:
 - Check normality → Non-parametric
 - Post-hoc tests

	sum_sq	df	F	PR(>F)	\
C(condition)	27.200476	1.0	30.577000	3.622839e-08	
C(SRC)	84.398095	4.0	23.718708	3.395379e-19	
C(HRC)	1119.378095	6.0	209.721969	1.813062e-208	
C(condition):C(SRC)	52.525714	4.0	14.761495	6.821003e-12	
C(condition):C(HRC)	7.536190	6.0	1.411949	2.061458e-01	
C(SRC):C(HRC)	56.688571	24.0	2.655233	2.314275e-05	
C(condition):C(SRC):C(HRC)	43.587619	24.0	2.041598	2.078639e-03	
Residual	1805.833333	2030.0	NaN	NaN	NaN
	eta_sq	omega_sq			
C(condition)	0.008508	0.008227			
C(SRC)	0.026398	0.025278			
C(HRC)	0.350118	0.348351			
C(condition):C(SRC)	0.016429	0.015312			
C(condition):C(HRC)	0.002357	0.000688			
C(SRC):C(HRC)	0.017731	0.011050			
C(condition):C(SRC):C(HRC)	0.013633	0.006954			
Residual	NaN	NaN			

Roma3 - DCR: 10s vs 20s - Vive Exploration



CWI - DCR: 20s vs 30s - Oculus Rift MOS



New scene features after 20s (motion)

CWI - DCR: 20s vs 30s - Oculus Rift



Main points

- **Preliminary** results from 3-way anova

- Significant effects:

- HRC
 - SRC
 - **Condition: to verify (SRC Vaude)**

- Interactions:

- Condition vs. SRC: Significant
 - **Condition vs. HRC: Non significant**
 - SRC vs. HRC: Significant

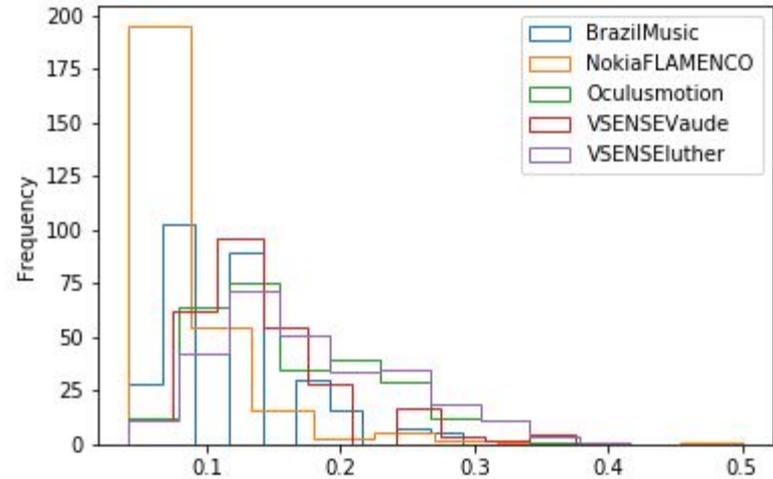
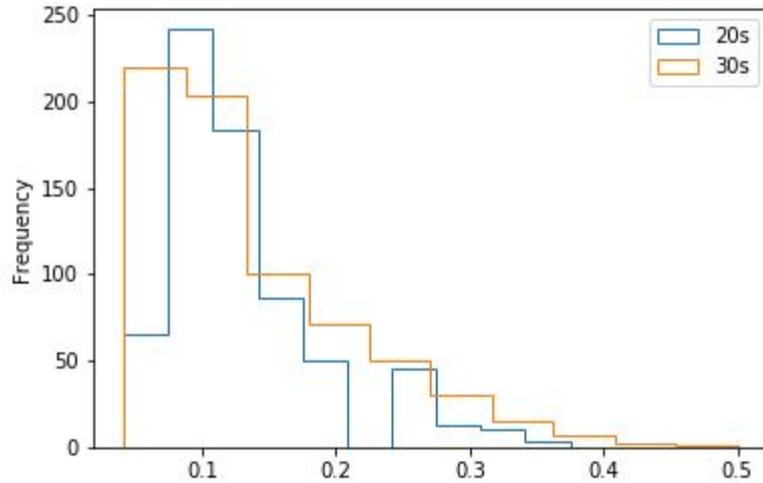
- Further analysis required:

- Check normality → Non-parametric
 - Post-hoc tests

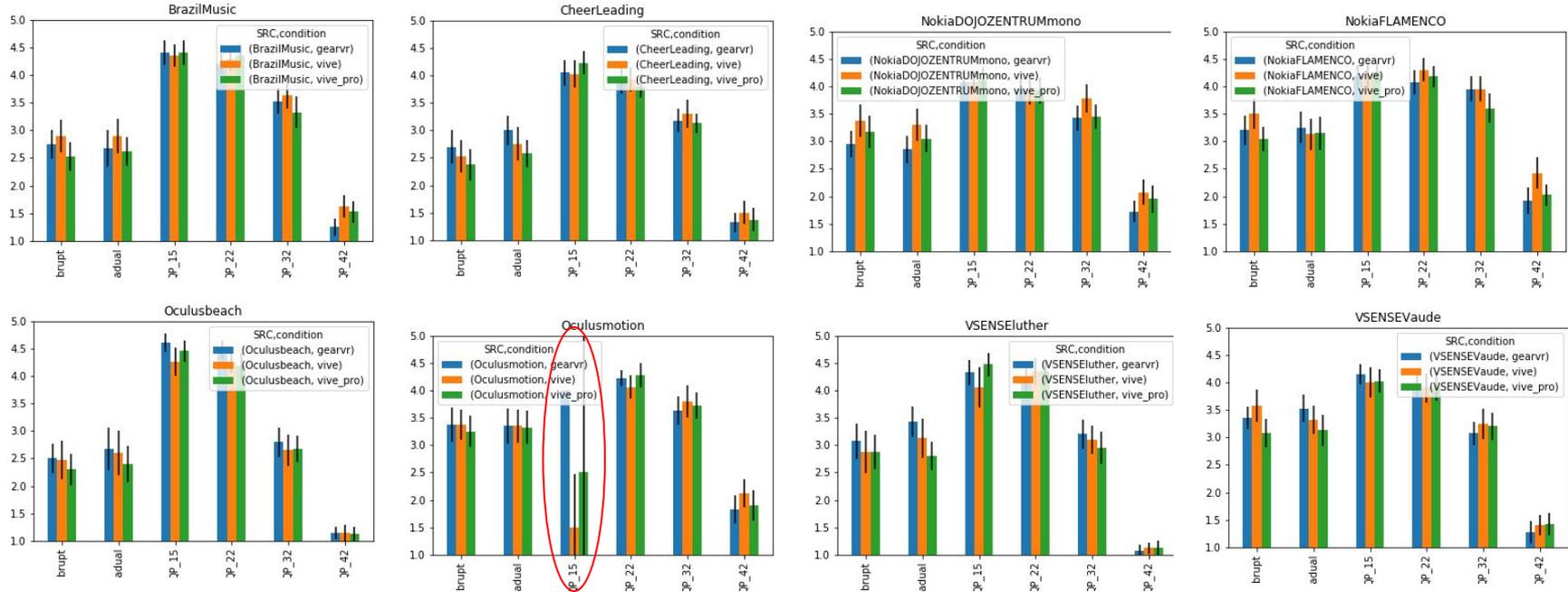
	sum_sq	df	F	PR(>F)
C(condition)	3.017857	1.0	5.559481	1.852285e-02
C(SRC)	58.602857	4.0	26.989473	1.428393e-21
C(HRC)	1526.660000	4.0	703.101369	0.000000e+00
C(condition):C(SRC)	11.042857	4.0	5.085774	4.546827e-04
C(condition):C(HRC)	1.385714	4.0	0.638189	6.353076e-01
C(SRC):C(HRC)	92.718571	16.0	10.675356	3.852751e-26
C(condition):C(SRC):C(HRC)	7.464286	16.0	0.859417	6.171898e-01
Residual	732.821429	1350.0	NaN	NaN

	eta_sq	omega_sq
C(condition)	0.001240	0.001017
C(SRC)	0.024080	0.023182
C(HRC)	0.627296	0.626265
C(condition):C(SRC)	0.004537	0.003644
C(condition):C(HRC)	0.000569	-0.000323
C(SRC):C(HRC)	0.038098	0.034521
C(condition):C(SRC):C(HRC)	0.003067	-0.000502
Residual	NaN	NaN

CWI - DCR: 20s vs 30s - Oculus Rift Exploration



UPM/Nokia - ACR20s - Gear vs. Vive vs. VivePro MOS



Problems with QP15 - OculusMotion
Not playing smoothly

UPM/Nokia - ACR20s - Gear vs. Vive vs. VivePro



Main points

- **Preliminary** results from 3-way anova
 - Significant effects:
 - HRC
 - SRC
 - **Condition: to verify (Problem with OculusMotion?)**
 - Interactions:
 - Condition vs. SRC: Significant
 - **Condition vs. HRC: Significant (Problem with OculusMotion?)**
 - SRC vs. HRC: Significant
- Further analysis required:
 - Check normality → Non-parametric
 - Post-hoc tests

```
"
```

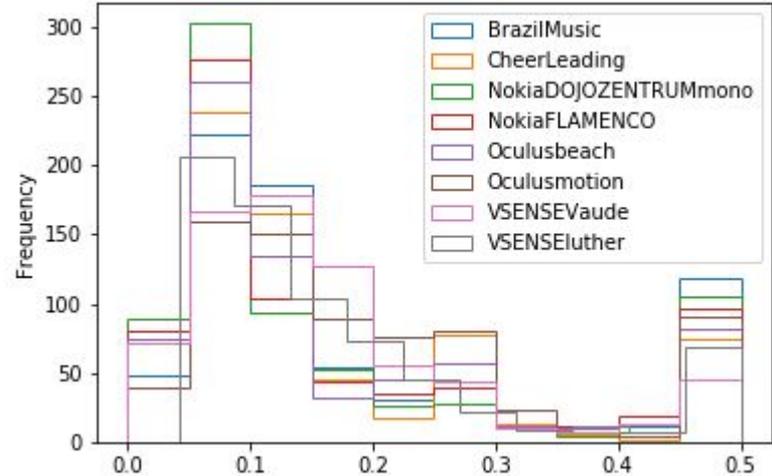
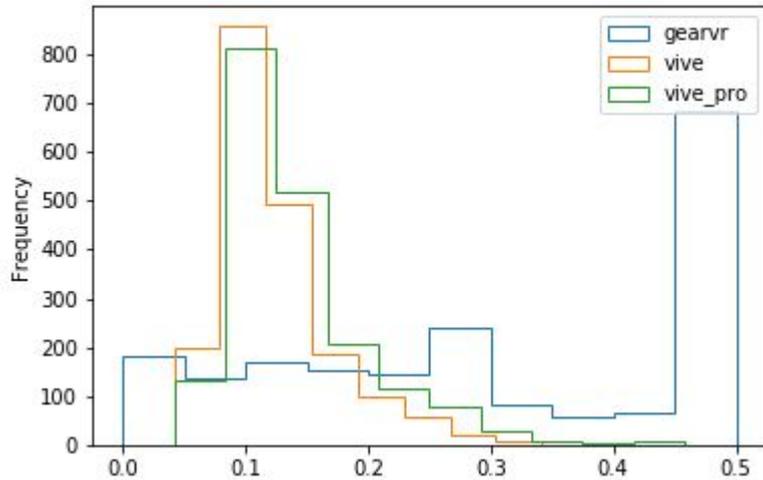
	sum_sq	df	F	PR(>F)	\
C(condition)	8.604514	2.0	6.509689	1.500202e-03	
C(SRC)	194.077983	7.0	41.950983	4.927692e-58	
C(HRC)	4491.466554	5.0	1359.195956	0.000000e+00	
C(condition):C(SRC)	19.005227	14.0	2.054040	1.142204e-02	
C(condition):C(HRC)	27.789980	10.0	4.204866	7.742648e-06	
C(SRC):C(HRC)	288.987292	35.0	12.493226	5.703951e-68	
C(condition):C(SRC):C(HRC)	34.830533	70.0	0.752880	9.380403e-01	
Residual	3711.617308	5616.0	NaN	NaN	

	eta_sq	omega_sq
C(condition)	0.000980	0.000830
C(SRC)	0.022114	0.021585
C(HRC)	0.511768	0.511353
C(condition):C(SRC)	0.002165	0.001111
C(condition):C(HRC)	0.003166	0.002413
C(SRC):C(HRC)	0.032928	0.030290
C(condition):C(SRC):C(HRC)	0.003969	-0.001303
Residual	NaN	NaN

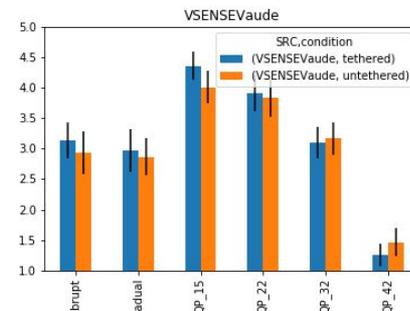
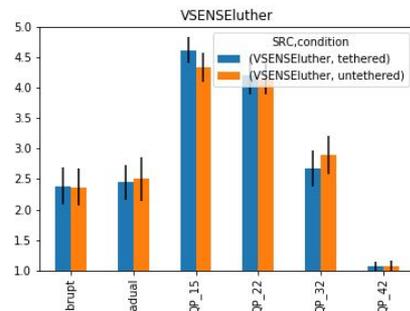
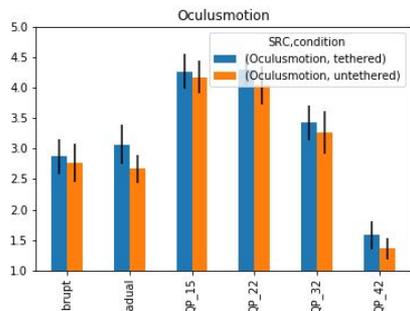
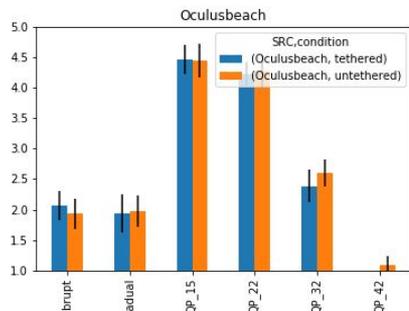
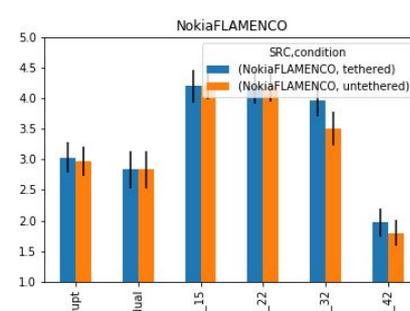
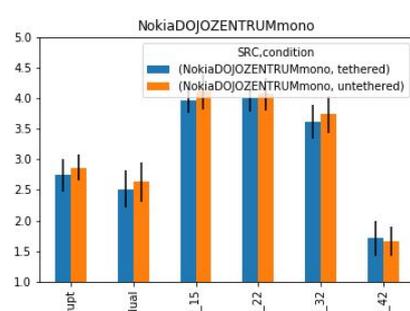
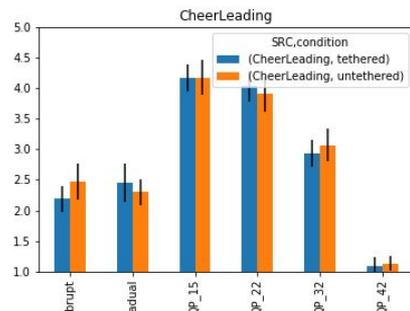
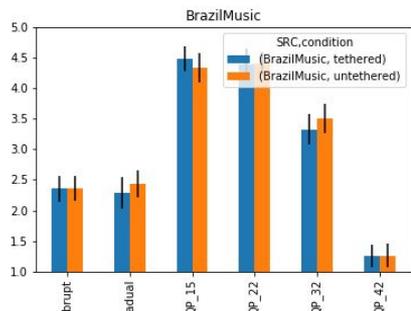
```
"
```

UPM/Nokia - ACR20s - Gear vs. Vive vs. VivePro Exploration

- **Almost no differences in exploration** (to further analyze)
 - We have to correct coordinates extracted from Gear!!!



Ghent - ACR20s - Tethered vs. Untethered - VivePro MOS



Ghent - ACR20s - Tethered vs. Untethered - VivePro



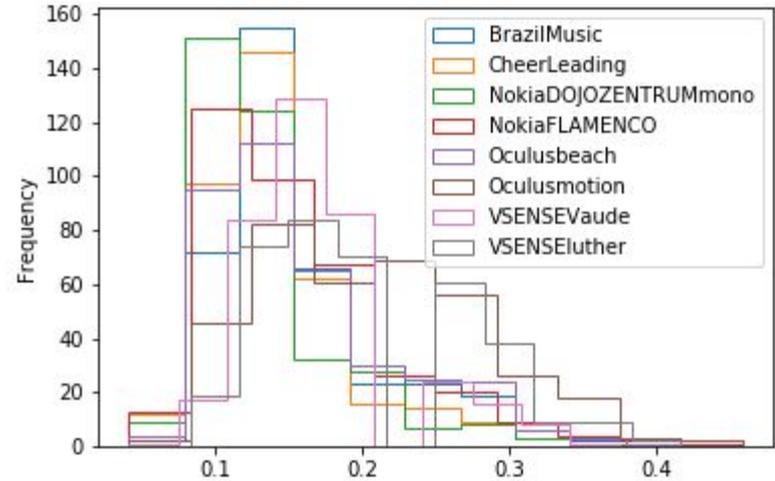
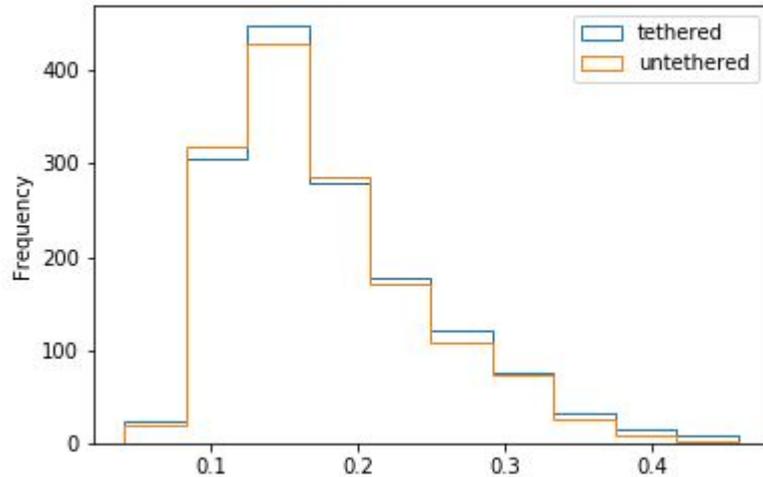
Main points

- **Preliminary** results from 3-way anova
 - Significant effects:
 - HRC
 - SRC
 - **Condition: Non significant**
 - Interactions:
 - Condition vs. SRC: **Non significant**
 - **Condition vs. HRC: Non significant!**
 - SRC vs. HRC: Significant
- Further analysis required:
 - Check normality → Non-parametric
 - Post-hoc tests

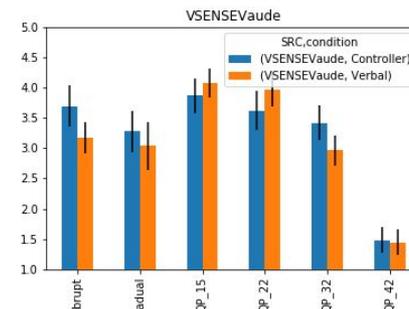
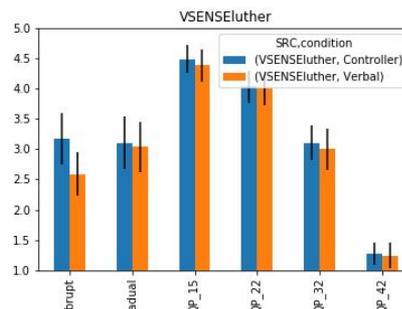
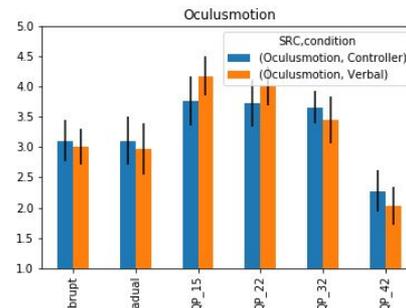
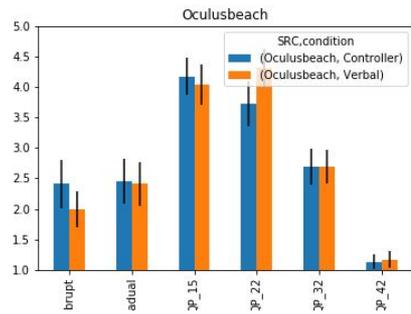
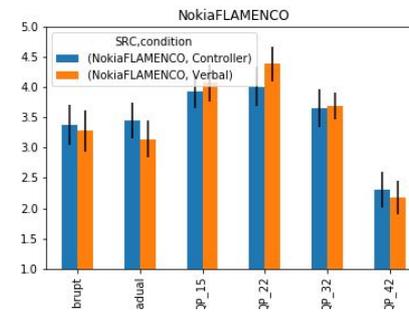
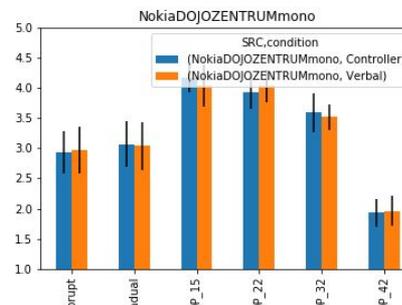
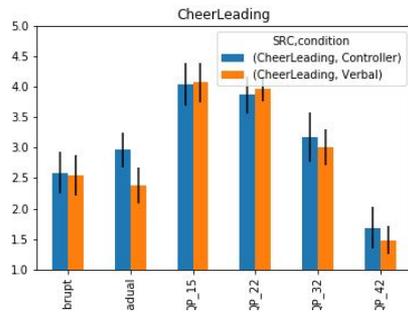
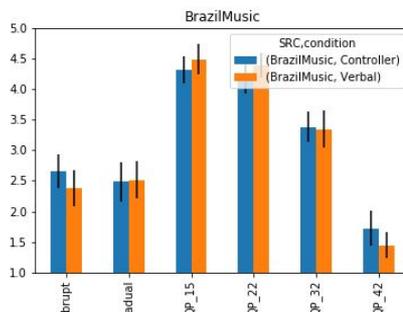
	sum_sq	df	F	PR(>F)	\
C(condition)	0.552806	1.0	1.045181	3.067069e-01	
C(SRC)	100.632046	7.0	27.180464	2.403160e-36	
C(HRC)	2908.805133	5.0	1099.925394	0.000000e+00	
C(condition):C(SRC)	5.931054	7.0	1.601963	1.300696e-01	
C(condition):C(HRC)	1.220945	5.0	0.461684	8.049882e-01	
C(SRC):C(HRC)	145.518820	35.0	7.860854	5.914472e-37	
C(condition):C(SRC):C(HRC)	11.757978	35.0	0.635160	9.531283e-01	
Residual	1496.813978	2830.0	NaN	NaN	
	eta_sq	omega_sq			
C(condition)	0.000118	0.000005			
C(SRC)	0.021543	0.020748			
C(HRC)	0.622706	0.622070			
C(condition):C(SRC)	0.001270	0.000477			
C(condition):C(HRC)	0.000261	-0.000305			
C(SRC):C(HRC)	0.031152	0.027186			
C(condition):C(SRC):C(HRC)	0.002517	-0.001446			
Residual	NaN	NaN			

Ghent - ACR20s - Tethered vs. Untethered - VivePro Exploration

- **Almost no differences in exploration with and without cables** (to further analyze)



TUI - ACR20s - Controller vs. Verbal voting - Vive MOS



TUI - ACR20s - Controller vs. Verbal voting - Vive

Main points



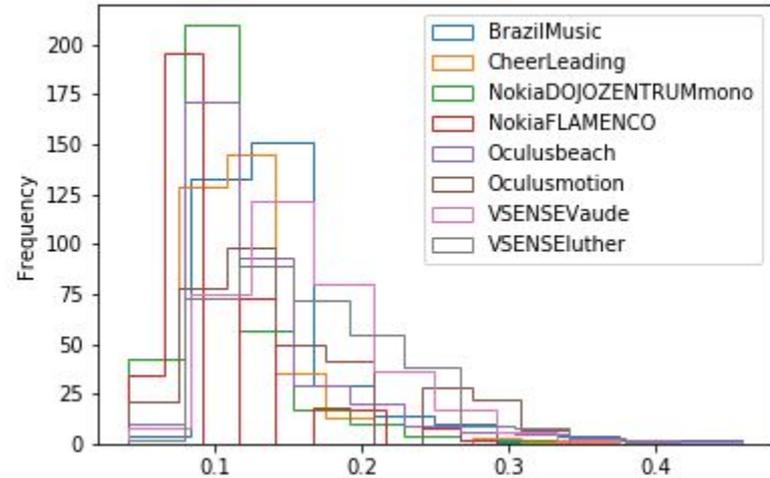
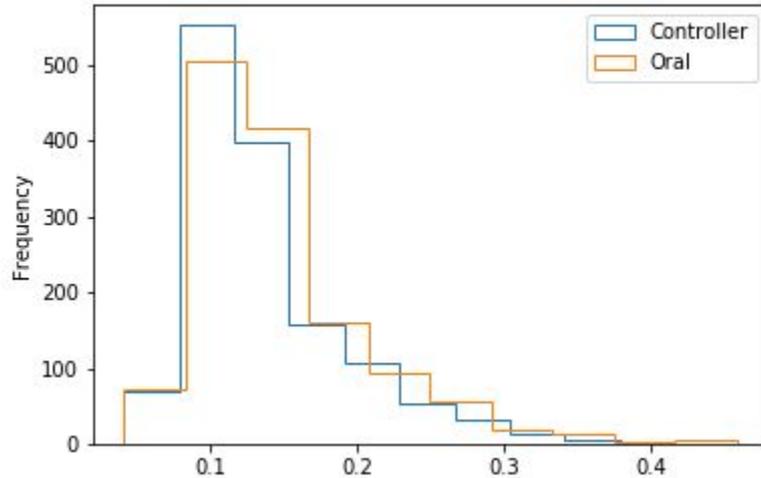
- **Preliminary** results from 3-way anova
 - Significant effects:
 - HRC
 - SRC
 - **Condition: Significant (to verify)**
 - Interactions:
 - Condition vs. SRC: **Non significant**
 - **Condition vs. HRC: Significant (to verify)**
 - SRC vs. HRC: Significant
- Further analysis required:
 - Check normality → Non-parametric
 - Post-hoc tests

```

"
sum_sq      df      F      PR(>F) \
C(condition)  2.185345  1.0    3.066311  8.004470e-02
C(SRC)       103.678161  7.0    20.781914  2.127982e-27
C(HRC)       1865.985632  5.0    523.642134  0.000000e+00
C(condition):C(SRC)  2.941092  7.0    0.589531  7.649935e-01
C(condition):C(HRC)  19.297414  5.0    5.415336  5.817288e-05
C(SRC):C(HRC)  142.485632  35.0   5.712146  6.306009e-24
C(condition):C(SRC):C(HRC)  18.507184  35.0   0.741940  8.649764e-01
Residual    1915.724138  2688.0  NaN      NaN

eta_sq  omega_sq
C(condition)  0.000537  0.000362
C(SRC)       0.025469  0.024239
C(HRC)       0.458383  0.457427
C(condition):C(SRC)  0.000722 -0.000503
C(condition):C(HRC)  0.004740  0.003864
C(SRC):C(HRC)  0.035002  0.028869
C(condition):C(SRC):C(HRC)  0.004546 -0.001581
Residual    NaN      NaN "
```

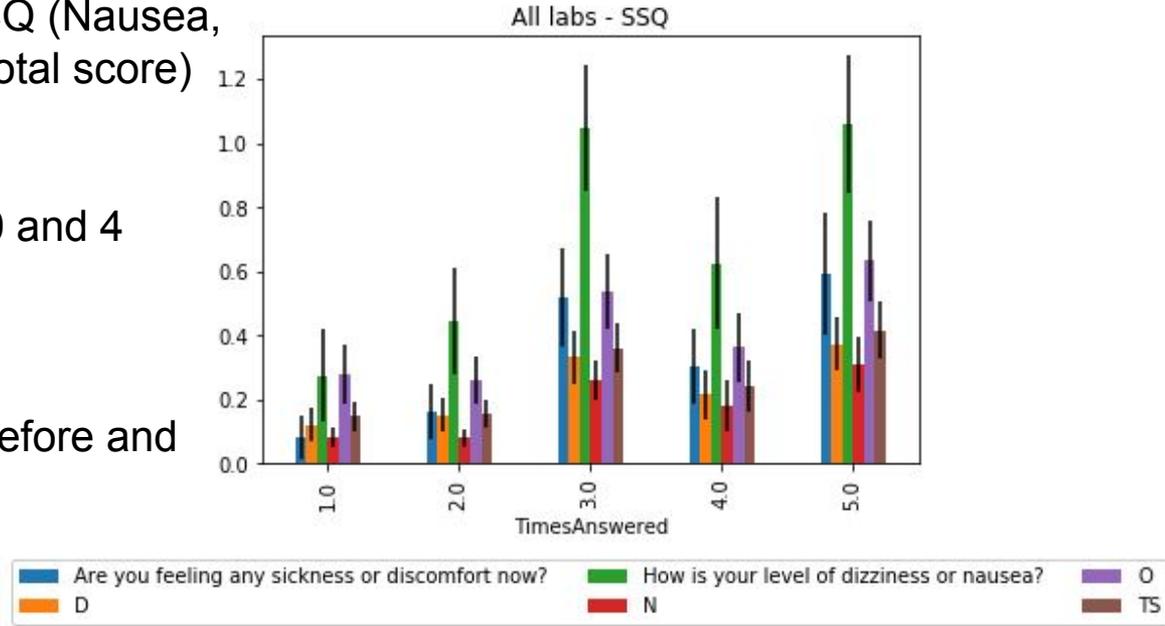
TUI - ACR20s - Controller vs. Verbal voting - Vive Exploration



SSQ

When to measure?

- Average scores for short SSQ questions and global scores for long SSQ (Nausea, Oculomotor, Disorientation, Total score) for all labs.
- Normalized values between 0 and 4
- Positive effect of the break
- No big differences between before and after the training.



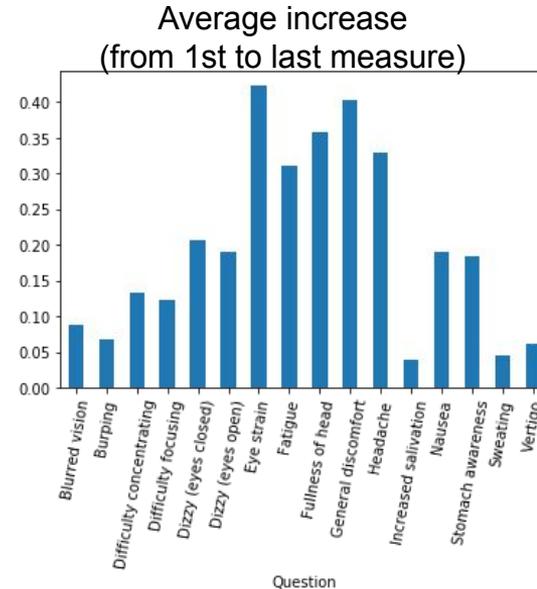
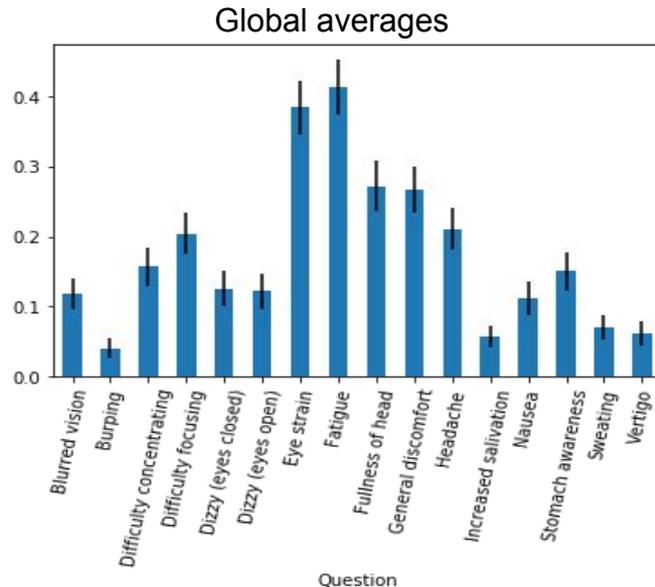
What to measure?

- Correlation among the different questions (average for all observers in each lab).
- Good correlation of short question “Are you feeling any sickness or discomfort now?” with global scores from the long SSQ.

Question	Are you feeling any sickness or discomfort now?	D	How is your level of dizziness or nausea?	N	O	TS
Are you feeling any sickness or discomfort now?	1.000000	0.880287	0.729281	0.928642	0.918340	0.944810
D	0.880287	1.000000	0.906602	0.812625	0.944481	0.965659
How is your level of dizziness or nausea?	0.729281	0.906602	1.000000	0.613290	0.790261	0.814928
N	0.928642	0.812625	0.613290	1.000000	0.876518	0.924910
O	0.918340	0.944481	0.790261	0.876518	1.000000	0.983559
TS	0.944810	0.965659	0.814928	0.924910	0.983559	1.000000

What to measure?

- Are all the factors from Kennedy's SSQ relevant for 360° video? Can we remove some items?
 - Most relevant factors seem to be: Eye strain, General discomfort, Fullness of head, fatigue, headache



Comments reported from labs

Main points

- Slight lip sync issues (Flamenco, Vaude)
- Issues with the app:
 - Some difficulties voting with the interface (RISE). Maybe a button to confirm the vote can help (AGH)
 - UPM and Ghent: Due to some issues, the test session with one subject was interrupted. They had to redo the whole session again. It would be nice to allow to re-start where the session was interrupted.
- DCR: people complain that position should not be reset between reference and PVS.
- Voting time was set to 5 seconds.
 - Some labs reported that it was too short.
 - Decision to disable timeout
 - Some labs used timeout to avoid using the controller to vote (Ghent).
- HW problems to play OculusMotion QP15 (AGH, UPM): bad MOS results due to them.
- Tests with long videos (30s): people complain of the test duration and the length of the sequences.

More to add...

Conclusions



Main objectives - Contribution to P.360-VR

Methodology:

1. Recommend:
 - a. ACR DCR for 360° videos: We have validated the methodologies used in the test. We can propose them as recommendation to ITU-T.
 - b. Sequence duration: We haven't found any significant difference between the sequences that we have tested. We can recommend 10 seconds for visual artifacts (uniform or non-uniform)
2. Study of influence factors:
 - a. HMD: Any commercial HMD (tethered or untethered) can be used, provided that it has enough resolution / frame rate to represent the content that is going to be tested.
 - b. Source content characteristics: To further explore and confirm.
 - c. Uniform vs non-uniform artifacts: To further explore and confirm.

SSQ:

1. Recommend:
 - a. One multi-item questionnaire (SSQ or derivation from it), and/or one single-question item
 - i. Short questionnaire A (vertigo scale) correlates better with Kennedy's SSQ → Proposal?
 - ii. Reduce number of items from Kennedy's SSQ
2. Write some guidelines about: More analyses are needed
 - a. When/how to assess simulator sickness
 - b. When can you use a single-question vs. full questionnaire
 - c. How to process the results / what to do with them

Ongoing steps

Deeper analysis of the obtained results

1. Post-hoc tests (+ non-parametric analyses?)
2. Observers' analysis: Outlier rejection, MLE?
3. Inter-lab analysis
 - ACR vs DCR
4. Content factors
 - Camera motion
 - With vs. without audio
5. Deeper exploration/navigation analysis
6. Deeper SSQ analysis
7. Background questionnaires

Exploitation of results

Proposal



1. Contribution for ITU-T P.360-VR

2. One journal paper
 - With the conclusions of the study and dataset
 - Authors: people from each participant lab which have contributed significantly to the work
 - Editors/writers (first authors): Jesús Gutiérrez & Pablo Pérez

3. Public data set
 - All test result files (raw data)
 - All PVSs (if possible) – Where? CDVL?
 - Linked to the journal paper (for citation)

Discussion



Questions? Comments?

- Thanks to:
 - All people working in the tests in all participant labs!
 - Providers of the source sequences

- For example:
 - How to proceed with the contribution for ITU-T P.360-VR
 - Comments/suggestions on the analysis of the results
 - How to proceed with the publication of the results and dataset

- More?

- Future steps:
 1. Phase 2: Long sequences (presence)
 2. Monoscopic vs stereo, Different resolutions, ...

Qualinet White Paper on Definitions of Immersive Media Experience

- Collaborative effort of QUALINET, the European Network on Quality of Experience in Multimedia Systems and Services.
 - QUALINET is known for the "Qualinet White Paper on Definitions of Quality of Experience" [2] and QUALINET Databases [3].
- White paper on Definitions of Immersive Media Experience → ideally < 10 pages
- Aim: to provide survey of definitions of immersion and presence which leads to a definition of immersive media experience (IMEx).
 - QoE for immersive media: relationship between the concepts of QoE and IMEx
 - Application areas of immersive media experience.
 - Influencing factors on immersive media experience
 - Assessment of immersive media experience.
 - Standardization activities related to IMEx are highlighted

Qualinet White Paper on Definitions of Immersive Media Experience

- First version already finish by the Qualinet community
- Now to be shared for community review at large:
 - Send it to standards/industry fora as well as research communities (MPEG, JPEG, VQEG, ITU-T, DASH-IF, VRIF, ACM SIGs, IEEE TCs, etc.)
 - Request feedback until April 1, 2020
 - Produce the final version to be presented at the QUALINET2020 meeting in Athlone (26-28 May 2020)
- Should we provide inputs as VQEG-IMG?
 - Coordination from chairs: collect interest, share document, set up audiocalls, ...
 - Should we leave it to individual contributions?

IMG Work plan Report

Phase 1 - Short Sequences

Jesús Gutiérrez

Grupo de Tratamiento de Imágenes (GTI)

Universidad Politécnica de Madrid (UPM)

jesus.gutierrez@upm.es

Pablo Pérez

Nokia Bell-Labs

Madrid, Spain