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

VQEG Plenary Meeting
12 March 2020
Immersive Media Group

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](mailto:img@vqeg.org)
Immersive Media Group

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.
Immersive Media Group
Quality assessment for 360° content

• Working on a test plan: https://docs.google.com/document/d/18L_Q7vyzEssoAzbRsTKBTtMrwfjISMf0HUGofPqOOJl/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 (<=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**.
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

<table>
<thead>
<tr>
<th>ID</th>
<th>Test condition</th>
<th>Lab</th>
<th>HMD1</th>
<th>HMD2</th>
<th>Comment</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>ACR: 10s vs 20s</td>
<td>Wuhan</td>
<td>Vive</td>
<td></td>
<td></td>
<td>Finished / 31 observers</td>
</tr>
<tr>
<td>B</td>
<td>ACR: 20s vs 30s</td>
<td>AGH</td>
<td>O. Rift</td>
<td></td>
<td></td>
<td>Ongoing</td>
</tr>
<tr>
<td>C</td>
<td>DCR: 10s vs 20s</td>
<td>Roma3</td>
<td>Vive</td>
<td></td>
<td></td>
<td>Finished / 30 observers</td>
</tr>
<tr>
<td>D</td>
<td>DCR: 20s vs 30s</td>
<td>CWI</td>
<td>O. Rift</td>
<td></td>
<td></td>
<td>Finished / 28 observers</td>
</tr>
<tr>
<td>E</td>
<td>HMD vs HMD (ACR 20s)</td>
<td>Nokia</td>
<td>GearVR</td>
<td>Vive Pro</td>
<td>Mobile vs desktop</td>
<td>Finished / 60 observers</td>
</tr>
<tr>
<td>F</td>
<td>HMD vs HMD (ACR 20s)</td>
<td>UPM</td>
<td>Vive</td>
<td>Vive Pro</td>
<td>Low res vs. High res</td>
<td>Finished / 40 per HMD</td>
</tr>
<tr>
<td>G</td>
<td>HMD vs HMD (ACR 20s)</td>
<td>Ghent</td>
<td>Vive Pro</td>
<td>Vive Pro</td>
<td>Tethered vs Untethered</td>
<td>Finished / 31 observers</td>
</tr>
<tr>
<td>H</td>
<td>With vs without audio (ACR 20s)</td>
<td>RISE</td>
<td>Vive</td>
<td></td>
<td></td>
<td>Ongoing / 18 observers</td>
</tr>
<tr>
<td>I</td>
<td>Scoring interface vs voice (ACR 20s)</td>
<td>TUI</td>
<td>Vive Pro</td>
<td></td>
<td></td>
<td>Finished / 29 observers</td>
</tr>
<tr>
<td>J</td>
<td>ACR: 10s vs 30s</td>
<td>Surrey</td>
<td>Vive</td>
<td></td>
<td></td>
<td>Finished / 31 observers</td>
</tr>
</tbody>
</table>
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...)

*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

<table>
<thead>
<tr>
<th>#Tiles</th>
<th>Transition</th>
<th>ROI</th>
<th>QPs</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>8x5**</td>
<td>Smooth</td>
<td>90°</td>
<td>42</td>
<td>37</td>
<td>32</td>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>6x3</td>
<td>Smooth</td>
<td>120°</td>
<td>42</td>
<td>32</td>
<td>22</td>
<td>22</td>
<td>32</td>
</tr>
<tr>
<td>8x5**</td>
<td>Abrupt</td>
<td>180°</td>
<td>42</td>
<td>42</td>
<td>22</td>
<td>22</td>
<td>42</td>
</tr>
<tr>
<td>6x3</td>
<td>Abrupt</td>
<td>120°</td>
<td>37</td>
<td>37</td>
<td>22</td>
<td>22</td>
<td>37</td>
</tr>
</tbody>
</table>
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
Sesion structure and duration

- Total session duration around 1h

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

<table>
<thead>
<tr>
<th>No problem</th>
<th>No perceptible effect, natural feeling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Light effects</td>
<td>Slight discomfort, but no sickness</td>
</tr>
<tr>
<td>Uncomfortable</td>
<td>Moderate discomfort, but tolerable for a while</td>
</tr>
<tr>
<td>Unpleasant</td>
<td>Strong discomfort or sickness, but can continue the test</td>
</tr>
<tr>
<td>Unbearable</td>
<td>Strong discomfort or sickness, and want to stop test</td>
</tr>
</tbody>
</table>

B. How is your level of dizziness or nausea?

<table>
<thead>
<tr>
<th>Very dizzy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dizzy</td>
</tr>
<tr>
<td>Slightly dizzy</td>
</tr>
<tr>
<td>Not dizzy</td>
</tr>
<tr>
<td>Absolutely not dizzy</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>ID</th>
<th>Test condition</th>
<th>Lab</th>
<th>Total</th>
<th>Female</th>
<th>Male</th>
<th>Age Min</th>
<th>Age Max</th>
<th>Age Avg</th>
<th>First time</th>
<th>Less than 5 times</th>
<th>5 to 20 times</th>
<th>More than 20 times</th>
<th>Every day</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>ACR: 10s vs 20s</td>
<td>Wuhan</td>
<td>30</td>
<td>15</td>
<td>15</td>
<td>20</td>
<td>30</td>
<td>24.5</td>
<td>8</td>
<td>15</td>
<td>7</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>ACR: 20s vs 30s</td>
<td>AGH</td>
<td>30</td>
<td>8</td>
<td>22</td>
<td>21</td>
<td>57</td>
<td>30.6</td>
<td>7</td>
<td>10</td>
<td>2</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td>DCR: 10s vs 20s</td>
<td>Roma3</td>
<td>28</td>
<td>14</td>
<td>14</td>
<td>21</td>
<td>60</td>
<td>27.6</td>
<td>2</td>
<td>12</td>
<td>5</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>D</td>
<td>DCR: 20s vs 30s</td>
<td>CWI</td>
<td>28</td>
<td>14</td>
<td>14</td>
<td>21</td>
<td>60</td>
<td>27.6</td>
<td>2</td>
<td>12</td>
<td>5</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>E</td>
<td>HMD vs HMD (ACR 20s)</td>
<td>Nokia</td>
<td>30</td>
<td>4</td>
<td>26</td>
<td>23</td>
<td>45</td>
<td>31.6</td>
<td>3</td>
<td>14</td>
<td>7</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>F</td>
<td>HMD vs HMD (ACR 20s)</td>
<td>UPM</td>
<td>30</td>
<td>4</td>
<td>26</td>
<td>23</td>
<td>45</td>
<td>31.6</td>
<td>3</td>
<td>14</td>
<td>7</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>G</td>
<td>HMD vs HMD (ACR 20s)</td>
<td>Ghent</td>
<td>30</td>
<td>4</td>
<td>26</td>
<td>23</td>
<td>45</td>
<td>31.6</td>
<td>3</td>
<td>14</td>
<td>7</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>H</td>
<td>With vs without audio (ACR 20s)</td>
<td>RISE</td>
<td>30</td>
<td>4</td>
<td>26</td>
<td>23</td>
<td>45</td>
<td>31.6</td>
<td>3</td>
<td>14</td>
<td>7</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>I</td>
<td>Scoring interface vs voice (ACR 20s)</td>
<td>TUI</td>
<td>29</td>
<td>14</td>
<td>15</td>
<td>20</td>
<td>37</td>
<td>25.9</td>
<td>13</td>
<td>12</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>J</td>
<td>ACR: 10s vs 30s</td>
<td>Surrey</td>
<td>31</td>
<td>10</td>
<td>21</td>
<td>19</td>
<td>44</td>
<td>25.9</td>
<td>13</td>
<td>12</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>238</td>
<td>90</td>
<td>148</td>
<td>19</td>
<td>60</td>
<td>27.04286</td>
<td>51</td>
<td>95</td>
<td>33</td>
<td>22</td>
<td>7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>Max</th>
<th>Avg</th>
<th>First time</th>
<th>Less than 5 times</th>
<th>5 to 20 times</th>
<th>More than 20 times</th>
<th>Every day</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>37.82</td>
<td>62.18</td>
<td>24.40</td>
<td>45.45</td>
<td>15.79</td>
<td>10.53</td>
<td>3.35</td>
<td></td>
</tr>
</tbody>
</table>
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
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)
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

<table>
<thead>
<tr>
<th></th>
<th>num_sq</th>
<th>df</th>
<th>F</th>
<th>PR (&gt; P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C(condition)</td>
<td>27.200476</td>
<td>1.0</td>
<td>10.877000</td>
<td>3.622839e-08</td>
</tr>
<tr>
<td>C(SRC)</td>
<td>84.398098</td>
<td>6.0</td>
<td>23.718708</td>
<td>3.358379e-19</td>
</tr>
<tr>
<td>C(HRC)</td>
<td>311.138045</td>
<td>6.0</td>
<td>209.721969</td>
<td>1.813062e-203</td>
</tr>
<tr>
<td>C(condition) : C(SRC)</td>
<td>52.525174</td>
<td>4.0</td>
<td>14.761495</td>
<td>6.521005e-12</td>
</tr>
<tr>
<td>C(condition) : C(HRC)</td>
<td>7.536190</td>
<td>6.0</td>
<td>1.441849</td>
<td>2.061950e-01</td>
</tr>
<tr>
<td>C(SRC) : C(HRC)</td>
<td>56.668571</td>
<td>24.0</td>
<td>2.515233</td>
<td>2.314278e-05</td>
</tr>
<tr>
<td>C(condition) : C(SRC) : C(HRC)</td>
<td>43.857619</td>
<td>24.0</td>
<td>2.041595</td>
<td>2.078639e-03</td>
</tr>
<tr>
<td>Residual</td>
<td>1805.833333</td>
<td>2030.0</td>
<td>NaN</td>
<td>NaN</td>
</tr>
</tbody>
</table>
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
CWI - DCR: 20s vs 30s - Oculus Rift Exploration
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
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

![Bar charts comparing MOS for different devices and conditions.](image_url)
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
Ghent - ACR20s - Tethered vs. Untethered - VivePro

- 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
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.

<table>
<thead>
<tr>
<th>Question</th>
<th>Are you feeling any sickness or discomfort now?</th>
<th>D</th>
<th>How is your level of dizziness or nausea?</th>
<th>N</th>
<th>O</th>
<th>TS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Are you feeling any sickness or discomfort now?</td>
<td>1.000000 0.880287</td>
<td>0.729281 0.928642</td>
<td>0.918340 0.944610</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>0.880287 1.000000</td>
<td>0.906602 0.812625</td>
<td>0.944481 0.965659</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How is your level of dizziness or nausea?</td>
<td>0.729281 0.906602</td>
<td>1.000000 0.613290</td>
<td>0.790261 0.814928</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>0.928642 0.812625</td>
<td>0.613290 1.000000</td>
<td>0.876518 0.924910</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>0.918340 0.944481</td>
<td>0.790261 0.876518</td>
<td>1.000000 0.983559</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TS</td>
<td>0.944610 0.965659</td>
<td>0.814928 0.924910</td>
<td>0.983559 1.000000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
SSQ
What to measure?

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

![Global averages](chart1)

![Average increase (from 1st to last measure)](chart2)
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.
  ○ Deicssion 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 360º 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)
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, ...

More?
Other Business

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.
Other Business

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

VQEG Plenary Meeting  
12 March 2020