VQEG

IMG Work Plan - what's next?

2021-06 – Virtual f2f meeting

Now what?

1. Further analysis of the data from the common experiment on 360 video
2. Complete “phase 2” of the test plan (long sequences)
   - ITU-T proposal: revise P.919 or make a new recommendation
3. Work on task-based and/or interactive use cases
   - As per original work plan
   - Explore new use cases
     - E.g. immersive collaboration
     - (COVID-19!)

Non-exclusive!

1. Cross-lab experiment: pending analysis

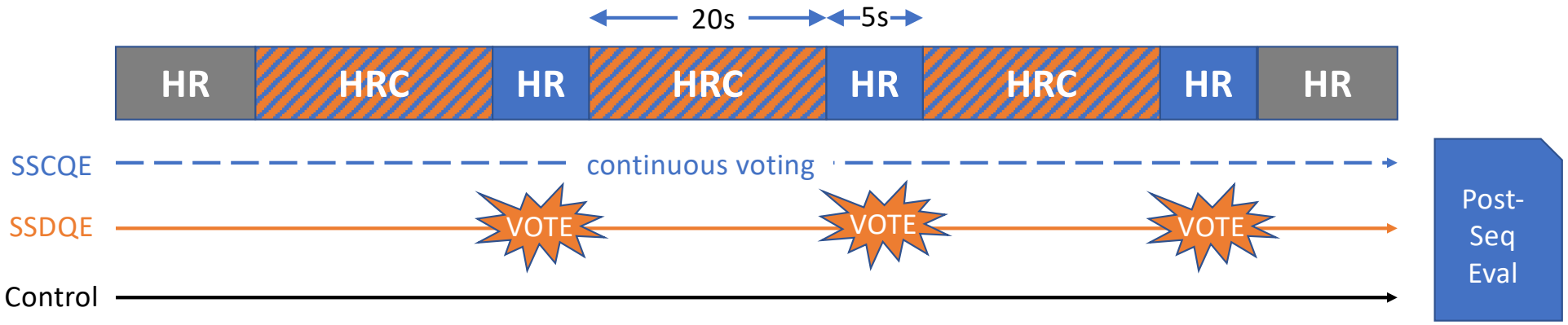
- What we have already:
  - Audiovisual quality: methodology, duration, HMD, audio, and rating method
  - SSQ: influence of methodology, long vs. single question and reduced SSQ
  - Exploration behavior: horizontal coverage
- What we are doing:
  - Subject bias / outlier detection
  - Exploration behavior
  - Lab-to-lab analyses
  - Computation of objective metrics
  - SSQ

2. Long Sequences
Introduction

- Definition: a short (2-10 minutes) **full** content item
  - Probably >90% of 360-degree relevant content
- Useful for *content immersive* evaluation [Pinson *et al*. 2014]
- In-sequence evaluation
  - Single-Stimulus Continuous Quality Evaluation
  - *Single-Stimulus Discrete Quality Evaluation* (SSDQE, aka CIETI) [Gutiérrez *et al.* 2011]
  - None! (control group)
- Post-sequence evaluation
  - Audiovisual Quality (ACR?)
  - Presence
  - Simulator Sickness
  - Task effectiveness

2. Long Sequences
Video Quality: what we have learnt (Orduna 2021)

- In-sequence evaluation
  - SSDQE provides “expected” results (similar to what ACR would provide)
  - Quality evaluation does not affect sense of presence
  - SSCQE → to be analyzed
- Post-sequence evaluation
  - ACR (whole sequence) is significantly affected by in-sequence evaluation
- Conclusion: SSDQE / SSCQE (?) / ACR are valid for long sequences
  - If you know what you are doing



2. Long Sequences
Other factors: what we know

- Simulator Sickness
  - Same physiological effect as with short sequences → no difference
- Presence (Spatial / Social)
  - It can be measured with ”traditional” tools (questionnaires)
  - No (or limited) interaction with quality / sickness measures
  - But… evaluation questionnaire very related with the content of the video
- Higher-level cognitive factors (empathy, attention, etc.)
  - Limited interaction with quality / sickness measures… as far as we know
  - But very difficult to standardize

2. Long Sequences
What can we propose to amend ITU-T P.919?

- Video quality for long sequences
  - Propose SSCQE / SSDQE for in-sequence evaluation
  - Propose ACR for post-sequence evaluation
  - Write some guidelines about the interaction between them
- Simulator sickness for long sequences
  - Use the same tool as for short sequences
- Presence and other cognitive factors
  - Write general guidelines, based on state-of-the-art, probably non-normative

Is this enough to complete the job? Should we do more experiments?

3. Task based / interactive use cases (summary from original work plan)

Use case		Free navigation	Semantic navigation	Task-based evaluation possible	Interactivity
Uni-directional	Entertainment	✓	+/-	✗	✗
	Training	✓	✓	✓	✗
Bidirectional	Machinery control	✓	v	✓	restricted (well-defined task)
	Human communication	✓	✓	✓	complex (free conversation)

3. Task based / interactive use cases
The problem

- Imagine you have a bi-directional immersive communication system
  - How do you test it?
    - Evaluate effect of technical factors in QoE (e.g. variations of latency / bitrate / etc.)
    - Compare with other systems / experiments
- ITU-T P.920 - Interactive test methods for audiovisual communications
  - Some tasks proposed to evaluate effect of technical factors:
    - E.g.: one of the subjects shows and describes a plastic building block and the other one is required to reproduce it;
  - Centered on video-conference (05/2000)
- ITU-T P.QXM - QoE Assessment of eXtended Reality (XR) Meetings
  - Best practices for QoE assessment of tele-meetings with extended reality elements
  - Work in progress (some VQEG members are contributors)

3. Task based / interactive use cases
Proposal of joint experiment

- Gather a set of **immersive communication systems**, e.g.
  - Real-time 360 video telepresence
  - Social VR with pointcloud transmission / with avatars
  - AR collaboration
- Create an experiment that covers **all basic functionalities**
  - Conversation between people
  - Discussion about objects in the immersive space
  - Interaction with (local / remote / virtual) objects in the immersive space
- Run a cross-lab experiment using **any available collaboration technology**
  - “The same” experiment in completely different setups

Target: Creating the “lego-block” experiment for VR/AR/XR

Now what?

1. Further analysis of the data from the common experiment
   - Discussion already started
2. Complete “phase 2” of the test plan (long sequences)
   - A proposal for video quality evaluation based on [Orduna 2021] experiment
     - **Is it ok? Should we do more tests? Address presence / immersion?**
     - **If yes, who would be interested in doing this??**
3. Work on task-based and/or interactive use cases
   - New joint work to **create single methodology for heterogeneous systems**
     - **Feedback?**
     - **Who would be interested?**

IMG

Reflector: img@vqeg.org

Jesús Gutiérrez: jesus.gutierrez@upm.es

Pablo Pérez: pablo.perez@nokia-bell-labs.com