 VQEG Meeting Minutes - Stockholm July 2014

# wendesday AM, July 7

## 3DTV

### Introduction by Barkowsky Marcus

Description of the current work and the meeting schedule.

1. Generating ground truth.
2. Frame compatible 3D video format representations.

### Test plan for VQeg grotruqoe3d1 database by Jing Li

For the ground truth experiment the sources have to meet specific conditions, like depth range within the comfort zone. It was shown that the relation between scores given to the quality question and the comfort question can be very different for different subjects.

Q: Did you have some additional questioner for each user so the source of those differences can be analysed?

A: Probably not, it has to be checked.

**Conclusion** (from the presentation not discussion)-> users do not have experience with 3D so ACR scale is not correct subjective experiment method. **Solution** -> pair comparison, the only question is which you prefer in terms or QoE and/or PoE. The problem of pair comparison is the number of pairs to present. For example, an ACR test taking 10 minutes changes to more than 350 minutes for pair comparison test. **Solution** -> better pair comparison focused on comparing specific pairs based on the rectangular design. A paper giving more explanation is linked in the presentation.

Q: Is the pre-test needed?

A: Yes, if the number of observers is small. If the number of observers is large the adaptation of the algorithm will convert to the correct solution.

Standard form for the data exchange was presented. All other information are provided by the test plan.

Q: The problem of the calibration was raised.

A: There is description of the calibration procedure.

Time schedule is proposed on slide 54 of the presentation.

### Test plan for VQeg grotruqoe3d1 database by Barkowsky Marcus

Each lab will get exact description which pairs should be compared by each observer.

Decision: Test sequences with the test plan are approved.

Q: Will we test if ACR is comparable with pair comparison?

A: After obtaining the grand truth data different subjective methods should be tested on the same PVSes so the comparison can be make.

Proposition: The size of display should be not smaller than 40 inch.

Florence Agboma: Market goes to the direction “bigger better” so 40 inch can be too little for the future use.

Chulhee Lee: The correlation between small size and large size is very high so it should not be important as long as the changes are visible.

Decision: Minimum display requirements: Full-HD resolution of the display panel (1920x1080 pixels), the recommended minimum display diagonal is: 40 inch‏.

Decision: Display technologies: Passive and active technologies may be used. Passive, polarized displays (notably line interleaved displays reducing the vertical resolution) are acceptable. Autostereoscopic displays may not be used if their resolution cannot be shown to be at least equal to passive polarized displays.

Decision: Viewing distance: A viewing distance of 3H will be used for active shutter glasses, for passive glasses displays with Full-HD resolution, a viewing distance of 4.5H will be used as a compromise between vertical and horizontal resolution (where H = Picture Height, picture is defined as the size of the video window, not the physical display.) For an (active or passive) Ultra-HD display a 3H viewing distance shall be used. Upscaling or Downscaling shall be done offline using a Lanczos-3 filter.

Decision: The 3D Display: Display shall be calibrated using a calibration tool before the test (see Annex L). The calibration must take into consideration the observer’s situation, i.e. glasses that are used during the experiment. If the result of the calibration is not satisfactory for an expert, factory settings may be used.‏

‎Decision: Background lighting: The exact value shall be reported. It is allowed to use no background lighting‏

Decision: Stimulus in between presentations: Preferred as is a 3D structured gray stimulus, for example gray squares at different depth, gray screen at Y=80 possible (to be documented) . This may also apply to the voting screen interface, i.e. a 3D voting screen may be preferred‏

Decision: Depth acuity test of subjects: (Randot Stereo test or equivalent )

Decision: Depending on the availability of the appropriate equipment, a subjective assessment with a Blu-Ray3D player with slightly compressed videos and a consumer type 3D display is acceptable because the sequences show a wide range of 3D quality conditions.

More details for the test condition and procedure can be found in the test plan.

The list of participants:

BSkyB : 2nd round, time sequential viewing, (eventually time parallel), 1 test

FuB: 2nd round, 24 observers only, time sequential, 4K display, 0.5 test

Yonsei: 1st round, 40 observers, time sequential, 55” polarized, 1-2 tests

NTIA: 2nd round, time sequential, 1 test

UPM: 1st round, time parallel, 1 test

INSA: 2nd round, time sequential, 1 test

T-Labs: 1st round, time sequential, 1 test‏

# Monday PM, July 7

## avhd

### presentation on AV experiments by margaret pinson

Summary of publication: “Audiovisual Quality Components: An analysis”. Different experiments come up with different conclusions. An experiment conducted by Margaret Pinson in 2010 was focused on proving which results are correct. The conducted experiment shows that probable explanation of obtaining different results is caused by using small number of source sequences in the experiment.

### Subjective test plan

Following discussion from yesterday.

Q: What is the current proposition?

A: Since the metrics are working on single component (audio or video) the experiment should refer to this.

Patrick Le Callet: The process should not be slowdown. Proposition: Run experiment asking only one question. We are using validated metrics therefore we do not have to validate it for video and audio separately. An AV test should be used to propose a common model and a future test will be run only on those PVSs where the proposed model does not work.

A: It needs knowledge about which video model and which audio model should be used.

Jari Korhonen : Content strongly influences what is more important, audio or video quality, for example football -> mostly video, music concert -> audio mostly.

Lip Synch excluded right now.

To avoid strong content dependency the content should be balanced by type of sound and video content. The taxonomy of content will be provided by Patrick Le Callet.

### h.265/hevc streaming evaluation by qi wang

It was remote presentation.

### ITU-t p.nats by Marie-Neige Garcia

The models have four modules, audio quality, video quality, combination both quality parameters, and the module taking into account starting time, quality switches, and stalling. Results should be ready by the next SG 12 meeting.

### vqeg/qualinet joint activity by Marie-Neige Garcia

Literature overview of the AVHD adaptive streaming. It is open for anyone. Already 50 papers with description. The output is the paper: “Quality of Experience and HTTP adaptive streaming: a review of subjective studies”. It is planned to run joined subjective experiments.

### Recent results on adaptive streaming by Brunnström Kjell

The test was run as crowdsourcing experiment. Pair comparison was used.

Q: How many subjects per point was used.

A: 15

Comment: Maybe it should be increased based on AV test conducted by Margaret Pinson.

Possible problems with the presented results, like better quality for 3 Mbit/s than for 5 Mbit/s, were pointed and discussed.

There was presentation of a tool to monitor a user connection.

### discussion about subjective testing methods for longer sequences

Christian Schmidmer pointed out that the problem of involvement in the content cannot be neglected if long sequence is considered.

VQEG should different from P.NATS. It could be differenced by considering not only parametric models but also non-parametric models, higher resolution, mobile devices, and short term MOS pooling function.

Chulhee Lee: We should remember that mobile phones are not necessary displaying Full HD.

James Goel: New use cases should be identified and described. Also we should consider extra layer generated by MPEG-DASH.

Vittorio Baroncini: HDR should be considered.

### Interest

Who is interested in AV quality integration (A and V to AV) function. Especially running subjective experiments.

University of West of Scotland

AGH

Ghent University – iMinds

Intel

DT

Qualcomm

British Sky Broadcasting Ltd

Technical University of Denmark

FUB

IRCCyN, Université de Nantes

## Announcement

### vqeg management

Arthur Webster is going to move to a different position within VQEG which calls for the next co-chair. Margaret Pinson was proposed and approved to be VQEG co-chair. In order to make the transition smooth Margaret Pinson become the third co-chair from this meeting.