VQEG Meeting in Los Gatos, CA USA

Hosted by Netflix

# Monday, May 8, 20017

* Liaison from SG12: “LS on study of G.QoE-VR", asks for responsibility of IMG, response positive.
* Liaison from SG12: “New ITU-T Recommendation in the work programme: G.QoE-VR",just information
* Liaison from SG12: “LS on continuation of joint activities under Intersector Rapporteurs Group on Audio-Visual Quality Assessment (IRG AVQA)”, asking SG9 about intention to stay in IRG-AVQA, to be discussed in detail during IRG-AVQA session, decided to leave this liaison without action
* New meeting host: AGH (PL)? Canon (Quan) (AU)? StreamOwl (Savvas) (GR)? Asia? UWS (Naeem), June 2019 (UK)?, Nokia (Pablo) (Spain)
* IMG (Immersive Media Group) session. Introduction by Patrick Le Callet. Changed order of presentations:
	+ Zhenzhong CHEN (Wuhan University)    IMG    Benchmarking VR Video Quality Assessment
	+ Pablo Perez Garcia (Nokia Bell-Labs)    IMG    Project Vertigo: monitoring sickness and discomfort in high-motion 360 video
	+ Glenn Van Wallendael (Ghent University - imec)    IMG    Subjective evaluation of 360 degree video quality during head movement
	+ Jesús Gutiérrez (University of Nantes)    IMG    A dataset of head and eye movements for 360 degree images
	+ Narciso García (Universidad Politécnica de Madrid, UPM))    IMG    360VR User Behavior
* QUALINET Presentation (Kjell Brunnström). Questions:
	+ What is the opinion of VQEG? Is there anyone opposing that VQEG continues this work? No. Chris Schmidmer (OPTICOM) highlighting that without clear projects the cooperation will work only w.r.t. sharing databases etc. Dale Stolitzka (Samsung) expressed positive feedback looking forward for more collaboration.
	+ Do we need the Term of Reference document? With the current scope it may not be needed - as highlighted by Chris Schmidmer (OPTICOM).
	+ What reflector to use. Currently there are two!
		- One hosted by VQEG.
		- One hosted by Qualinet. Rather this one, as the software behind it simply works better.
	+ Co-chair, project-leaders from VQEG?
		- Possibly IMG could co-chair, proposal by Patrick Le Callet, decision postponed to Thursday morning
		- Maybe there could be joint VQEG-Qualinet chairs for similar projects, proposal by Chris Schmidmer
* Other presentations:
	+ C.-C. Jay Kuo (University of Southern California)    AVHD    How to Link Perceptual Visual Coding and Quality Assessment?
	+ Lukas Krasula (UN)    HDR    Methodology for Objective Metrics Performance Evaluation...

# Tuesday May 9, 2017

Note taker: Christian Schmidmer, OPTICOM GmbH

## PsyPhyQA

### Project overview presentation given by Sebastian Bosse (VQEG\_VLQA\_2016\_118\_\*)

### Project overview presentation given by Sebastian Bosse (VQEG\_VLQA\_2016\_119\_\*)

         Presentation of AGHs test facilities where some experiments were conducted.

         PsyPhyQA is looking for 3..5 min long source sequences which fulfil their requirements (see Sebastian's slides)

         Ioannis pointed to the Chimera video which contains longer sections without scene cuts

         Mikolaj pointed out, that he can run experiments for interested parties at very low (no?=) cost since this is part of the multimedia course curriculum. Interested parties shall contact him directly.

### Presentation by Naeem Ramzan, "Perceptual Video Quality Evaluation by Means of physiological Signals" (VQEG\_VLQA\_2016\_125\_\*)

Will be presented at QOMEX

Description of experimental protocol and setup for EEG, ECG and EMG

Sequences are split ("classified") into two sets, high and low quality.

The following methods were investigated to let the machine predict the classification based on physiological data:

|  |  |
| --- | --- |
| EEG | 56 features |
| ECG | 84 features |
| EMG | 21 features |

Different classifiers were examined (1-NN, 3NNm SVM (linear) SVM (RBF kernel))

Best results were achieved with "Feature Selection" and only 6 features.

Comments by Patrick:

It may also be interesting to look at the results of the Qualinet task force lead by Kathrien De Moor. This task force has also investigated the comparison between consumer and medical grade EEG devices for QoE studies.

As EMG signals possibly captured by EPOC have not been filtered out, classification on EMG signals only seem very promising.

### Presentation by Sebastian Bosse, "Objective quality assessment of stereoscopic images with vertical disparity using EEG" (VQEG\_VLQA\_2016\_119\_\*)

## VIME

Michele: Introduction to VIME and how INTEL was using the VIQET tools

### Presentation, Lucjan Janowski,  "Interpreting VIQET Scores, When can a user see a difference?"

Subjects made a pair comparison of images presented side by side using two identical monitors.

Images were not identical frames, but same scene, shot with different cameras

Subjects had to choose which of the two was better.

Differences between results obtained by AGH and the other labs may be due to problems with the display calibration at AGH.

A formula could be derived which describes which (VIQET-)MOS difference can be perceived by a certain percentage of users.

Ranking according to MOS created by crowd sourcing (one monitor, one image at a time) was not always correlated to rank order in side-by-side lab test. Conclusion is that scoring MOS and selection based on pair comparison are not necessarily the same thing!

### Presentation Mikolaj Lesczuck, "Automatic extraction of Machine Tags in Flickr Service"

## IMG

### Presentation Ashutosh Singla, "QoE of Omnidirectional (360°) Videos"

In most cases HTC Vive was preferred for 4k content

Simulator Sickness: Strong correlation between sickness an QoE

Occumulator = Similar to average of the nausea indicators

There are still open questions as to why subjects moved their heads so little.

### Presentation Suiyi Ling, "3 Datasets for Quality Assessment in the Context of free Viewpoint Video"

### Presentation Jesus Gutierrez, "Characterization and selection of light field content for perceptual assessment"

Subjective and objective methods for quality assessment are needed

Both require proper content

Various 2D and 3D dimensions are used to characterize scenes in the content space

## HDR / WCG

### Presentation on progress by Patrick

### Presentation Jesus Gutierrez, "Integration of VQM on VMAF Framework"

Main focus is extension from 8 bits to 10 bits in order to integrate with the VMAF framework

### Presentation Lukas Krasula, "HDR/WCG Content Characterization Method"

Characteristics are proposed which can be used to select HDR content which covers a representative range for e.g. subjective testing.

## UltraHD

### Presentation, Jesus Gutierrez, "VMAF framework performance on UHD videos"

Motivation is the ICIP Grand Challenge which requires objective metrics to preselect compression algorithms which outperform H.265.

## eLetter

Introduction to next issue by Jesus. Topic of the issue: immersive Media. Interested parties can still contribute!

Proposed topics of the future issue:

         AVHD-AS / P.NATS Phase 2 (probably too early, to be discussed in AVHD sessions)

         PsyPhyQA (tentatively accepted)

         Haptic Feedback (should go into immersive media issue)

Session closed

### Presentation, Werner Robitza, "Behavior-oriented Quality of Experience Assessment in the lab and at home"

Subjects were given the chance to interrupt a viewing session when they do not tolerate the degradations anymore. Reasons for stopping were Initial loading, stalling and compression artifacts. Subjects reacted very different. Test design is critical and ends in dilemma between influencing users and low accuracy / few data points. Furthermore, a crowdsourcing extension for YouTube is presented which can collect information on user behavior from real YouTube sessions.

There was some concern by the streaming service providers in the room regarding such investigations. Concerns ranged from privacy issues to violations of net neutrality or interfering with the service providers quality switching control loop.

Comment by Ioannis: there is a service recently acquired by Netflix which can be used safely and without spam etc. to analyze the network connection: fast.com.

## General

Lucjan reminded the group of the open access "Quality and User Experience" journal and encourages contributions (<http://www.springer.com/engineering/signals/journal/41233>).

Naeem bids for QOMEX 2019 and proposes a colocation with a VQEG meeting at UWS. He also asks for people being interested in joining the program committee etc.

# Minutes

# May 10th, VQEG meeting

# AVHD-AS / P.NATS

## #24. Florence Agboma (Sky Broadcasting). Service provider requirements for adaptive streaming.

Presentation:

* “Who we are”. Entertainment (PayTV) & Telecom service provider.
* Purpose: provide the requirements from a typical service provider.
	+ There’s no monitoring product currently in the industry that can give QoE metrics for adaptive streaming.
* Typical end-to-end VoD workflow (see slides for the diagram)
	+ There are QoE analytic tools in client devices (NPAW, Streamlyzer, etc), but not (sufficiently) in the rest of the chain.
* Service provider requirements for VoD content delivery
	+ At each component in the workflow, there must be possible to measure and monitor quality checks.
* Final thoughts.
	+ Linear content is similar to VoD in distribution side (differences: short-lived CDN, manifest refresh).

Q&A (some of them)

* Q. What is mezzanine format?
* A. Intermediate quality format (e.g. 25 Mbps) used as input into transcoders.
* Q. What can we do with the metrics once we have them?
* A. Discussion about that. It can be used to relate to customer engagement. Lead to conversations with network providers. Negotiate capacity with CDN providers.

## #7. Christian Schmidmer (OPTICOM). Curing the Video Playout Nightmare.

Presentation

* Motivation. Some subjects reported sporadic flicker in videos. Difficult to track: is it in the video, in the player, in the test setup?
* Develop a flicker sensor (based on USB soundcard + photosensor) and added alternate white/black patches to the video.
* Test results: some players (e.g. VLC), there were sporadic flickers about every 10 seconds
* Moved to a new test setup, with full control of the signal path (see slide for details):
	+ Blackmagic Design DeckLink 4k card + Gen2 PCIe slot (8 lanes)
	+ 2x NVMe SSD.
* Measures with the new card were much better (quite accurate).
* Used the tool to detect actual frame rate in the screen (and detect frame interpolation in the screens, e.g. 3:2 pulldown effects).
* Also, used to measure LCD TV backlight (pulsed), OLED TV white level…

Q&A.

* Some concerns about the accuracy of the sensor. Questions left for the coffee break.

## Jörgen Gustafsson & Silvio Borer. AVHD-AS / P.NATS agenda of the interim meeting.

* Officially an ITU meeting.
* Goal: Standardized method to measure video quality for adaptive streaming services over a reliable delivery (no packet loss).
* Phase 2: bitstream modes, pixel models, hybrid modes.
* (See schedule in documentation)

# VIME

## #23. Pini Zorea (ORT Braude Engineering College). From prediction of perceived image quality to online image quality improvement.

Presentation:

* The idea is transitioning from image evaluation to image improvement (e.g. For social networks).
* Phase 1: relationship between standard IQAs and perceived IQ
	+ Take images -> modify IQAs parameters (brightness, etc.) in 10% steps -> DSIS subjective assessment -> MOS
	+ 98 non-expert students (80% male, 20% female), 5 per session
	+ Mapping between “standard” image quality attributes and VIQET image quality categories.
* Phase 2: Introducing VIQET parameter extraction in the loop. Calculating MOS based on VIQET parameters and specific coefficients found during the test.
* Phase 3: Subjective assessment (comparing Human Visual Test mode with MOS predictions).

Q&A:

* Discussion about how MOS provided by VIQET was not used.

# VLQA

(Basically a re-run of the Special Session on Visually Lossless Video Quality @ HVEI 2017).

## #10. Kjell Brunnström (Acreo). Overview of HVEI special session on Visually Lossless Video Quality for Modern Devices: Research and Industry Perspectives

* Presentation of the session held on Human Vision and Electronic Imaging in Burlingame CA, USA, Jan 29 – Feb 2, 2017.
* Target: cover visually lossless compression from a quality assessment point of view.

## #13. Laurie Wilcox. Subjective assessment and the criteria for visually lossless compression

Presentation:

* Subjective assessment of the effectiveness of VESA Display Stream Compression (DSC1.2) standard. 120 test subjects.
* Tested using ISO/IEC flicker protocol (29170-2:2011 Annex B). Which image is flickering (in 4 seconds, side-by-side comparison). One image is compressed (DSC) alternated with its original (5 Hz flicker rate), while the other is original-original.
* Criteria for visually lossless: all observers fail to correctly identify the reference on >75% of the trials.
* Naïve observers, trained to find artifacts as part of the test session.
* Statistical detection of outliers (they are critical as a single outlier can break the all-below-75% criteria).

NOTE: Correct schedule excel sheet. It was supposed to be given by James Goel.

## #14. Phil Corriveau. Gaze-contingent flicker paradigm.

Presentation:

* Joint Research Venture: Intel + Pacific University Oregon
* Lossless images: human vision is limited by visual acuity and attentional resources. There’s a need for a new test paradigm (flicker methodology).
* Research on methodology
	+ 30 adults (18-35 year old), >= 20/25 binocular acuity
	+ Eye tracker
	+ Conditions: original-original, original-compressed, original-compressed after fixation onset, original-blurred
	+ Blurred: Gaussian mask over the diff image (original – compressed). Image is blurred where compression artifacts happen.
* Gaze-contingent image switching: image is only switched just when user fixes gaze in the image (min 150ms for saccadic motion to happen).
* Conclusion: standard methodology (flicker) probably overestimates detection of lossy images, by generating “artificial” attention to the artifacts (which is good by design).

NOTE: Correct presentation title in the schedule.

## #17. Al Ahumada (NASA). Spatial-Temporal Visible Contrast Energy Predictions of Detection Thresholds

Presentation:

(Lots of equations and graphs – better see presentation pdf to get an idea!)

* Spatial Model
* Spatial-Temporal Model
* Spatial-Temporal Detection Threshold Data
* Parameter Estimation and Model Fit

Q&A:

* Flickering enhances visibility of low-spatial-frequency artifacts, and therefore it could over-estimate the visibility of low frequency components of the artifacts.

## #34. Glenn Van Wallendael (Ghent University - imec). Visually lossless video quality evaluation methodology

Presentation:

* Focused on video (not still images).
* Experiment 1. Single-Stimulus Video. Select videos on the high-end (MOS >=4) and get users to vote whether they have seen the difference or not.
	+ Which question to ask?
* Experiment 2. Side-by-side video playback, choose which one is better.

Q&A:

* Long discussion about the validity of the test and the possibility of detecting JNDs properly (and how it can be improved).

# QART

## #3. Mikołaj Leszczuk and Lucjan Janowski (AGH). Evaluation of Video Summarisation

Presentation:

* Target: summarize a video and translate into another language.
* How can be evaluate the “correctness” of a video summarization?
* Elements to consider: “artistic” considerations (video summary looks ok), all parts of message are there, length (shorter is better), audio+video issues, completeness of information.

Q&A (some ideas of the conversation)

* Do not use MOS. Maybe a single overall question about artistic rating is OK.
* For “correctness” analysis, have some of the users watch the whole video and generate some “ground truth tags” that can be compared with their interpretation of the video.
* Generate test questions about the video (similar to SAT tests).
* Consider TRECVID.

## #5. Margaret Pinson (NTIA/ITS). Public safety update

Presentation:

* PSCR Prizes & Challennges.

Discussion about viewing distance:

* What is the recommended viewing distance to be able to see the content correctly? Example: a jury needs to be able to watch a video that is used as proof for the trial in a court room. How far can they be from the screen?
* Some ideas:
	+ Military standards
	+ Acuity tests (information from medical standards)
	+ Check ITU-R BT.500
	+ Show a test pattern to the jury and verify they can see it properly

Decision: Margaret will lead the work below

* Propose a simple recommendation about viewing distances that can be used in court rooms.
	+ Start from article written in eLetter.
	+ Draft a recommendation to discuss in next meeting.
	+ Generate a test pattern to use for “calibration” in court room. Mikołaj can provide. Phil will check with Pacific University.

# Presentations / AVHD

## #2. Margaret Pinson (NTIA/ITS) and Mikołaj Leszczuk (AGH). Subjective testing of 4 sec videos without reuse.

Presentation

* A metric that works at the beginning of the processing chain (quality of the source video).
* Design subjective assessment tests that can measure both original and processed video qualities.
* How do people judge video quality? -> can we divide MOS into factors?
* Subjective tests based on scenes (4 seconds): never show the same exact original content twice to the same user.
* ACR plus: computer error (system is broken), human error (e.g. you were distracted).
* Positive feedback from testers, even about duration (2 – 2.5 hours).
* Cover factors such as: blurred background, fast camera panning, … that don’t usually get covered in VQEG tests.

Q&A

* Why watching each sequence just once? 1) To avoid memory effects, and 2) to cover more different content (there are much more content sources than actual processing chains in real world).

# VLQA

## #12. Assistant Prof. Damon Chandler (Shizuoka University). Masked detection of HEVC compression artifacts on professional, consumer, and mobile displays

Presentation

* Can thresholds measured on mobile devices yield the same results as in professional monitors? How the variabilities compare to other factors?
* Professional screen, consumer, iPad. Calibrate to (roughly) same transfer functions. No statistically significant difference among displays.
* Different “brightness” settings on iPad (0%, 50%, 100%, 50% + daily room). No significant differences either.
* Different content, however, can lead to more different than monitor or settings.

Thursday

Meeting Minutes VQEG on Thursday 11/05/2017

Potential next VQEG meeting host

1. Mikolaj, AGH, Poland
2. Savvas, StreamOwl, Athens, Greece
3. Pablo, Nokia, Madrid, Spain

Imp: Nov. 27 to Dec. 1 is the most appropriate week for the next meeting. Maybe not a good week for Nokia, but still interested in hosting in the future. One other week was also identified as possible, but as not as good: Nov 6 to 10

Arifumi Matsumoto presented “QAPI: a QoE control technology for online video streaming services”

-------------------------

 (AVHD-AS/P.NATS) group is looking for source sequences

* 4K UHD1
	+ Short sequences test (no audio)
	+ Long sequences 1-5 mins with audio

Spirent can potentially provide the sequences (offline discussion)

Florence from Sky can potentially provide sequences.

--------------------------

Immersive Video Group

Discussion on future steps

* 360 systems
	+ Subjective assessment methodology
	+ Objective metrics
	+ Datasets and tools (Google can potentially share some sequences)
* Virtual reality, mixed reality, augmented reality
* Light field technologies, Free Viewpoint Video

----------------------------------

“Business perspectives on perceptually lossless and lossy quality” is presented by Scott Daly (Dolby)

----------------------------

Ioannis Katsavounidis (NETFLIX) presented “Dynamic Optimizer: Dynamic optimizer: a perceptual optimization technique for video coding”

--------------------

Afternoon Session

**JEG-Hybrid**

Glenn presented the vision of “JEG-Hybrid” project and discusses about current progress and future directions

Enrico Masala presented “Enlarging the JEG-Hybrid large database of video sequences”

Ahmed Aldahdooh presented “From large scale to small scale database”

Immersive Media Group short discussion

Kjell Burnnstrom presented “Balancing Type I Errors and Statistical Power in Video Quality Assessment”

Samira Tavakoli presented “Status update on ITU-T Crowdsourcing Standardization”

Friday

Presentation from Farhad Abbassi (Cisco), “Liaision between VQEG and IEEE video subgroup”

**Want better collaboration with VQEG. IEEE CPIQ is interested in developing a quality ruler / JND mapping (no-reference analysis). If interested, email** fabbassi@cisco.com **This will feed into the development of a standard metric that consumers can use to understand mobile cameras—a 5-star scale.**