

# **VQEG meeting minutes**

## **Berlin, Germany June 22 - 26, 2009**

Including Minutes from each day's sessions.

Note: the ITU-T JRG-MMQA meeting is held coincident with VQEG during the Multimedia and Hybrid sessions.

### **VQEG Minutes from Monday 22 June 2009, Berlin**

Thanks to Filippo Speranza (FR) of CRC who volunteered to be note-taker.

Introduction of participants.

Meeting logistics by Alexander Raake (AR).

Arthur Webster (AW) informs that the latest version of the RRNR-TV test report has been uploaded to the VQEG ftp.

### **Summaries of Project Status**

Kjell Brunnström (KB) presents the summary update of the ILG activity, mostly about HDTV work.

Chulhee Lee (CL) presents the updated of the RRNR report, which now includes the analyses requested at the last meeting.

Chris Schmidmer (CS) presents the update of the MM2 project. Discussion are still undergoing about methods. Some presentations will be given during this meeting.

Margaret Pinson (MP) presents the update of the HDTV. Model submission has been delayed to let ILG time to complete its activities.

Yves Dhondt (YD) presents the update of the Tools and Subjective Labs Setup Group. Limited progress.

CL presents the update of the Hybrid project. Most of the test plan is completed, but there are some issues to be discussed.

KB presents the update for the Project for Joint Effort Group (JEG). KB presents the agenda for Wednesday when the work of the JEG will be discussed

### **Liaison Reports**

IEC T100 – AW presents IEC interest for VQEG.

ITU-R WP6C – CL note that WP6C would like to be informed more promptly about VQEG work.

ITU-T SG12 – Akira Takahashi (AT) and Jörgen Gustafsson (JG) say that there are two questions (Q13 and Q14) for which VQEG activity, particularly for the Hybrid project, might be of relevance.

ITU-T SG9 – MP notes that the rapporteur group in SG9 is working on two possible RECs: one using the results of the RRNR-TV and one related to new PSNR calculation. AW reports some editorial work on J.247 and J.246.

ATIS IIF QoSM – AW notes that ATIS has sent a liaison, which will be discussed during the meeting.

## **HDTV**

MP says there are three issues:

1. New deadlines. Model submission = September 8. Tentatively, first version submitted by proponents = August 18.
2. Text to accompany video sequences & subjective data.
3. How to decide if common set sequence scores are similar enough to be valid.

MP presents the new Test schedule for the HDTV.

The model submission has been delayed to September 8<sup>th</sup>. Earlier versions should be submitted by August 18<sup>th</sup>. The new dates are voted. No objection.

CL and MP discuss how to share the video material. Proponents will send a new 2T hard drive before August 15<sup>th</sup> to NTIA-ITS. (Proponents might purchase the drive in the USA and have it delivered to NTIA). These drives will be used to share video sequences to other proponents.

We now discuss the possibility of submitting MOS obtained with alternate monitor (NTT proposal). The new date for this (optional) possibility is November 13<sup>th</sup>

The ILG can decide on any PSV that might need to be discarded. The new date is changed to October 29<sup>th</sup>.

The time available for the proponents to run the model on all subjective dataset is changed to October 29<sup>th</sup>.

The date within which the objective score are validated is changed to November 27<sup>th</sup>.

The date within which the ILG fit the objective model data is changed to December 11<sup>th</sup>.

The date for proponents to submit alternate fit is also changed to December 25<sup>th</sup>.

Statistical analysis should be completed by January 28<sup>th</sup>.

The draft final report should be ready by February 27, 2010.

Approval of the final report March 27, 2010.

The new schedule is as follows:

1	Approval of test plan.	January 27, 2009
2	ILG issues an estimate of cost to participate in HDTV Test, based on feedback recorded at the San Jose meeting.	February 11, 2009
3	Date to declare intent to participate, the number of models that will be submitted.  All proponents who will participate in the HDTV test must specify their intent by this date.	February 17, 2009
4	Proponents supplied SRC made available to all proponents and ILG	March 22, 2009
5	ILG post monitor specifications to the HDTV Reflector.	As soon as possible, to allow replacement. February 26, 2009
6	ILG wanting to use purchased SRC obtain agreement from other ILG and Proponents.	March 8, 2009
7	ILG identifies fee for each proponent, and gives the proponent an invoice. ILG and proponents agree on a payment date.	March 3, 2009
8	Fee payment due. Proponents with special needs may negotiate a different deadline.	March 31, 2009
9	Sample video sequences distributed to ensure program interface compatibility.  Chulhee Lee will create some test vectors.	February 28, 2009
	Proponents send a new 2TB hard drive to NTIA/ITS. This hard drive will be used to send the video sequences to proponent. To save on shipping costs, proponents are encouraged to purchase the hard drive in the US. NTIA/ITS will send out an email identifying some US companies where hard drives can be purchased.	August 15, 2009
10	Proponents submit the first version of their model	August 18, 2009
11	Proponents submit their models to ILG.	September 8, 2009
12	Video sequences and subjective data distributed to all ILG and Proponents.	September 22, 2009
13	[Optional] proponents submit MOS for experiments using an alternate monitor (see section 2.2).	November 13, 2009
14	ILG decides on any PVSs that may need to be discarded.	October 29, 2009
15	Objective model data run on all subjective datasets.	October 29, 2009
16	Objective scores checked (validated).	November 27, 2009
17	ILG fit objective model data to subjective data.	December 11, 2009
18	Proponents optionally submit replacement model fit coefficients	December 25, 2009
19	Statistical analysis	January 28, 2010
20	Draft final report.	February 27, 2010
21	Approval of final report.	March 27, 2010

22	Subjective data published (all experiments)	Released with the HDTV Final Report
23	Objective data published (only models in the Final Report)	The following ITU-T SG9 or ITU-R SG6 meeting
24	Video sequences made public (only experiments to be made public)	Released with the HDTV Final Report

The text of HDTV test plan which discuss the use of VQEG - HDTV subjective rated datasets is examined.

CL suggests providing only a subset (e.g. 80%) of the data. FS suggests providing the data only upon signing a document in which the organization requiring the data agrees to certain conditions. Quan Huynh-Thu (QT) thinks that once the datasets are made public, we can only rely on the ethical behaviour of the requesting party. We have legal standing to prevent misuse. CS said that we had already agreed on this issue. Nothing is changed. CL asks if the video is copyrighted. MP says it is not. Only limit is use for research purposes. MP says she will go back and edit the text to specify that the sixth dataset will be available only to VQEG ILG and proponents only. Paul Coverdale suggests further edits to explain that according to VQEG comparison of models using only the five public dataset would be misleading.

**Final Proposed Text:**

*These five HDTV subjectively rated datasets were created by the Video Quality Experts Group (VQEG) in 2009. These and an additional sixth dataset were designed to independently analyze HDTV objective video quality metrics submitted to VQEG. The official data analyses are available in the HDTV Final Report. The video sequences associated with the sixth dataset are available to VQEG ILG and proponents, only.*

*These HDTV subjectively rated datasets may be used for research and development purposes, only. Models that are trained on these datasets should not be compared to the models submitted to VQEG for independent validation in 2009. Such a comparison would be misleading, because the experiments contain mainly source scenes and HRCs that were unknown to the model developers. Additionally, this comparison would be misleading because the sixth dataset has been kept private.*

**This text is approved.**

We discuss the issue of -how to decide if common set sequence scores are similar enough to be valid-. No proposal is advanced. Accordingly, the ILG will analyse the data and report on possible ways of doing that. ILG will decide then.

**RRNR-TV**

CL presents the new version of the RRNR-TV (version 1.7) Final Report, which is now available on the meeting files section. This new version contains several new graphs illustrating the performance of the models for transmission errors and by codecs. In addition, Appendix B (SUBJECTIVE TESTING FACILITIES) the description of the FUB lab is still missing.

MP suggests some explanatory text to accompany the graphs. CL asks if the inclusion of these graph warrant modifying the summary. A sentence indicating the presence of these graphs in Appendix C is added at page 4 of the report.

The RRNR-TV is deemed completed. **The report is approved.**

## **HYBRID**

Morning. CL introduces the issues that will have to be discussed. The most critical issue is the system to generate the HRC.

MP asks whether this system is limited to the one decoder, the JM, currently employed. CL thinks that the system can use other decoders as long as the TS can be read by the JM decoder.

Silvio Borer (SB) note that the JM decoder might not work well with transmission errors.

JG recalls that Ericsson had proposed an alternate system (an offline system).

Afternoon. CL presents the revised version of the Hybrid test plan. This new version is available on the FTP site. CL proceeds to illustrating editorial modifications which implements decisions and changes **already approved at the last meeting.** The list of decisions is contained in the .ppt file named -hybrid\_models\_for\_discussion3.ppt.

In relation to one of these decisions, namely the viewing distance for SD format, David Hands (DH) asks why the distance decided is 4H and not 6H. A discussion ensues. It is suggested to switch to 6H consistently with the table at page 4 of Rec.500-11. **The change is approved.** The test plan is changed accordingly where appropriate.

The phrase in Section 4.1.10. Instructions for Evaluators that suggests paying the viewers to motivate them is deemed irrelevant and therefore it is being deleted. **All agree.**

Lucjan Janowski (LJ) asked about the full matrix approach. The issue still needs to be discussed.

It is suggested to change the duration of the QVGA source material in Section 6.1.1- Duration of Source Sequences. The duration is changed from 16 to 15 seconds and from 20 to 19 seconds.

**The change is approved.**

MP raises the issue of rewinding in Section 6.3.4. Pausing with Skipping and Pausing without Skipping. A new text is suggested “

*The video should not play backwards, because this is an unnatural impairment. However, the video may jump backwards in time in response to a transmission error, or display a portion of a previous frame along with the current frame.*

AR suggests not using decoders with such behaviour. MP says that similar errors can be introduced by encoders as well.

After a lengthy discussion, **the text is approved** under the condition that substantial changes introduced by different decoders will be addressed for bitstream models somewhere else in the test plan.

The editorial modifications to the test plan which implements decisions and changes already approved at the last meeting are completed. Revised test plan v1.5 is approved. The discussion now moves to the issues that still wait for a decision.

Issue 1: *Are we going to use numerical values attached to each category?*

The scale in B.2/P.910 will be used. The scale will be treated as an integer (discrete) scale.

The descriptions at the extremes will be modified to eliminate any reference to “faithful to the original”.

# VQEG Official Notes

## Hybrid Bit-Stream Test Plan Tuesday June 23, 2009

Thanks to Margaret Pinson (NTIA) for taking notes.

Discussion continues on “0” and “10” on the 11-point scale, debating the suitability of using the labels and end points (e.g., “no further improvement is possible” for 10).

Agreement was reached that subjects will not be told to avoid “0” and “10” on the 11-point scale.

Agreement was reached that either computer or paper entry can be used for subjective testing. (In either case, 11-point scale will be discrete / integers as decided Monday.)

Agreement was reached that when SDTV is displayed on an HDTV LCD monitor, the picture area should be centered and the non-picture area should be black or mid-level grey. (Note: this addresses a conflict in the test plan.)

Agreement was reached to have a common set containing 30 video sequences. These 30 video sequences will be included in the total number of sequences within each test (agreed to previously).

→ Keishiro from NTT presented document 20090622NTT\_01.pdf. The proposal was to merge the Hybrid NR method with the Bitstream 2 method into a single study item.

Proposal: Regardless, merge Hybrid NR method (uses actual PVS) and Bitstream 2 method (decodes bit-stream within model, no access to actual PVS). Or perhaps group them together in descriptions and presentation, to aid the reader.

Agreement was reached that for Hybrid models, multiple decoders/players will be used to generate PVSs, as long as the decoders can handle the bit-stream data which the reference decoder can decode. Bit-stream data can be generated by any encoder as long as the reference decoder can decode the bit-stream data.

→ Alexander from DT presented document COM12\_LS\_015.doc, liaison statement from ITU-T sg12 in response to the Hybrid project’s request for information on P.NBAMS and P.NAMS.

→ Source content availability – unless otherwise specified, the standard VQEG license agreement must be signed. The process of getting license agreements signed should be started soon.

Opticom – Yes

NTT – Yes

KDDI – 3 years

NTIA/ITS – Yes, no license needed; HDTV and public safety content also

Psytechnics – Yes

SwissQual – Yes

FT – unknown

Yonsei -- Yes

Acreo – Yes, no license needed

→ Jun from NTT presented document 20090622NTT\_02.pdf. NTT proposes that VQEG adds another hybrid perceptual model, which can be used for estimating video quality using packet

headers and video signals, to the standardization target of the hybrid perceptual/bitstream project (e.g., when video is encrypted).

Agreement: NTT proposal was considered and a decision was postponed till later.

→ Jens from Swissqual presented document hybrid\_models\_for\_discussion3.pdf. See that document.

→ Alexander from DT presented Temporary Document 22 rev1 (WP 2/12) from ITU-T SG12.

**Note:** Thursday morning will be two breakout sessions: JEG and Hybrid Test Plan.

The Hybrid test plan session may attempt to clarify the introduction, as a concern has been raised that the goal of this test is now unclear. We discussed writing a Terms of Reference and Scope for the Hybrid project.

# Wednesday June 24, 2009

## **Notetaker**

Thanks to Yves Dhondt of Ghent University for taking notes.

## **Review of the notes of Tuesday**

The notes were updated to reflect that no decision was made with regards to the NTT proposal on the use of encrypted video streams.

## **Joint Effort Group**

Introduction by Patrick LeCallet.

The presentations given today will be made available online at the VQEG FTP-site if they aren't yet.

**Presentation 1:** A No Reference (NR) and Reduced Reference (RR) Metric for Detecting Dropped Video Frames (NTIA – Margaret Pinson)

A MATLAB implementation of the algorithm and a paper describing it in more detail are available at [http://www.its.bldrdoc.gov/pub/ntia-rpt/09-456/09\\_456.pdf](http://www.its.bldrdoc.gov/pub/ntia-rpt/09-456/09_456.pdf)

The methodology is covered by several US patents but NTIA allows free usage.

**Presentation 2:** IPTV-interface open source software (Acreo – Kjell Brunnström)

The software is freely available from <http://www.acreo.se/iptvinterface>

There are some questions regarding the use of a different transport mechanism (e.g. RTP instead of MPEG TS) and the interface. As the software is using DirectShow, things should be easy to change.

**Presentation 3:** Task-Based Video Quality Update (NTIA – Carolyn Ford)

**Presentation 4:** xStreamer – Modular Multimedia Streaming (IBBT – Ghent University – Nicolas Staelens)

The software is freely available from <http://xstreamer.atlantis.ugent.be>

There is some discussion about the extensibility of the software and the use of multiple slices. As the application uses a graph structure, components should be easy to be switched out and replaced.

**Presentation 5:** AGH and Lancaster University (AGH – Lucjan Janowski)

There is some discussion about the results presented for blockiness, flickering, and fps.

**Presentation 6:** Considerations for video scene selection (NTIA – Margaret Pinson)

The presentation brings up several points regarding VQEGs past and current content selection and the possible issues with it.

**Presentation 7:** Study of Rating Scales for Subjective Video Quality Assessment Using Single-Stimulus Methods (Psytechnics – Quan Huynh-Thu)

Discussions on the results presented.

**Presentation 8:** Comparison of subjective test methodologies (University of Nantes – Patrick LeCallet)

Discussions on the results presented.

**Presentation 9:** On Confidence and Response Times of Human Observers in Subjective Image Quality Assessment (Blekinge Institute of Technology – Ulrich Engelke)

# Meeting minutes Thursday 25th June

Notes by Quan

Morning

JEG and Hybrid break-out sessions running in parallel  
JEG session

Last words « call for proponents » changed for « call for participation ».  
Could be a call for proponent later, but too early right now.  
Talk regarding the licensing. Should be discussed in parallel/online, but should be discussed.

Scope

Margaret: SRCs available, mainly 1080p 30. If wait until HD test, will have more available. SD mainly in NTSC 30.

Filippo: If exe available, could be distributed to people who have databases, and get the results, with some description of the test sequence.

Margaret: HDTV test material not available before early 2010.

No dataset (with bitstream) available at this stage. Marcus can generate HRCs in case some lab can run a test.

Smaller format easier to process.

NTIA has HDMI to RGB converter.

Small format: 7 organisations

HD: 3 organisations

All formats: 5 organisations

In any case, tools should be written in such a way that should work for different resolution.

Margaret: need at least one interlaced format, to avoid bad surprises later on.

Scope: is working at NAL level good enough ? What information do we lose by doing that ?

Marcus:

- packet info data (not available in bitstream, timestamps...)

- bitstream

- PVS

- SRC

Marcus : should ask at what information is needed by all model so that we do not a test that will be useless because one information was not collected.

Nicolas: when making PVS, should save all the information available (PCAP file...)

Patrick : Need action point to list what's need as input to model to deal with transmission errors.

Margaret: if start with no transmission error test (coding error only), allows to do fast simple test.

Margaret : should start with two small tests, so to check that all modules can handle changes in the 3 variables (resolution, progressive / interlaced, framerate)

Margaret: CDvial website, all members are encouraged to shoot video and upload.

Ugent: Maybe better to start with complete decoder, not NAL parser, because will need more information with time. Reference software.

Marcus: would like to see a vmware image with all tools to produce HRCs

Patrick: on HRCs, coding error will always be useful in the future, but requirements for transmission error may change.

Write the specs for the test, get it approved by the group.

Patrick: Dataset available to the group, with PVS, bitstream, MOS score. So metrics can be tested against it starting from now.

Filippo: One fast way to test the metric is to test them against full ref models.

Action points:

- ~ SRC collection, Margaret, Lucjan
- ~ subjective methodology studies, Ulrich, Filippo, Lucjan, Roland,
- ~ H264 NAL parser, UGent, Ircyn
- ~ metric framework (inputs...) to handle transmission errors, Marie-Neige, Marcus, UGent
- ~ software issue, Roland, Marcus, Margaret, Ugent, Lucjan, Kjell
- ~ first core experiment: available database metric
- ~ PVS generation =>different encoder (collect PVS generators, report on them), Marcus,
- ~ Ugent
- ~ decoders =>SoA and collection, Patrick
- ~ next call for evidence, Margaret (test)

Afternoon

Report of the discussions in the JEG session:

- Scope was discussed
- Actions points were identified
- Owners of some of the action points were identified

(see above)

Report of the discussions in the Hybrid session:

- Terms of reference for the hybrid models were defined in document VQEG\_Hybrid\_TOR rev2.doc
- The document still needs further editorial work
- Will be included as an Annex in the Hybrid Test Plan

Potential locations and dates for next VQEG meeting:

- Chile in January 11
- Japan in January/February

LS from ATIS IIF about QoE survey

- There was some concern about how/where ATIS is going to report/disclose the information provided in the answers of the survey
- Will ATIS share the information with other organizations, e.g. VQEG?

Email from Gartner to VQEG

- AW informed the group about an email from Gartner requesting information about telepresence

#### MM2 session

- Before the meeting, the chairs had sent out an email requesting any interested party to present their experience in running audio-video subjective tests
- Presentations were provided by DT labs, NTIA/ITS, Symmetricom and Orange Labs
- CS also presented an overview of existing speech and audio subjective test methods
- Discussions:
  - NTIA mentioned that there is
    - a need for audio models applicable to a wide range of content as currently models are only for speech or music
    - a need for a wide variety of content in validation of models as relationship between audio and video qualities may be highly content-dependent
  - Can NTIA share the sequences used in their studies so that another lab can repeat the experiments to examine stability of audio-video quality assessment tests?
  - Should audio-video tests follow the recommendations for video tests (e.g. ITU-T P.910 or ITU-R BT.500) or recommendations for audio tests (e.g. ITU-R BS.1116 or MUSHRA or ITU-T P.800) or else (new method)?
  - Existing recommendations for audio quality assessment target mostly high-quality audio (very small impairments) using expert listeners for historical reasons (e.g. development of MP3)
  - Is there an important influence of the speakers or headphones on the stability/repeatability of the results?
  - Need to identify a methodology that can provide stable audio-video video quality scores across test labs
  - How can we replicate an audio-video test in different countries since the audio track may contain portions of speech content in a given language?

**Friday June 26, 2009**

**Note Taker**

Thanks to Stefan Winkler of Symmetricom for taking notes.

**Review of Thursday Notes**

The notes from the previous day were reviewed and approved.

**MM Phase 2**

Quan presents slides on MM2 project plan.

Proposal to conduct collaborative studies of subjective AV experiments regarding methods, setup, etc. Issues: language, audio reference, audio playback, audio level, audio quality range

Open request for further comments and suggestions on proposal.

**3DTV**

Markus presents slides on new ITU-R Question 128/6, other 3DTV issues (display technology, content characterization, etc.)

**Poll:**

- Who has 3D video system available? 2 parties (Yonsei, CRC)
- Who plans to buy one? 2 additional parties
- Who plans to develop 3D algorithms? 3 parties

Discussion of possible subjective experiments and source video content.

**Hybrid**

Decision to form working group to develop and qualify working system & reference decoder.

Candidate system to be reported at next meeting. Members: SwissQual, Ericsson, DT and Yonsei

**VQEG Document**

Proposal by Arthur to draft document on VQEG history and working methods, Wikipedia entry.

Participants: Arthur, Chris, Filippos, Magaret, Chulhee

**Next VQEG Meeting**

Likely to be in Santiago, Chile, in the week of January 11-15, 2010.

**Outgoing Liaisons**

Liaison to SG12, Report of JRG.

— Meeting ends at 12 noon —

[Telekom Labs tour after lunch]

## **Notes of the Joint Effort Group (JEG) breakout session**

Thursday, AM

Last words « call for proponents » changed for « call for participation ».

Could be a call for proponent later, but too early right now.

Talk regarding the licensing. Should be discussed in parallel/online, but should be discussed.

– Scope

Margaret: SRCs available, mainly 1080p 30. If wait until HD test, will have more available. SD mainly in NTSC 30.

Filippo: If exe available, could be distributed to people who have databases, and get the results, with some description of the test sequence.

Margaret: HDTV test material not available before early 2010.

No dataset (with bitstream) available at this stage. Marcus can generate HRCs in case some lab can run a test.

Smaller format easier to process.

NTIA has HDMI to RGB converter.

Small format: 7 organisations

HD: 3 organisations

All formats: 5 organisations

In any case, tools should be written in such a way that should work for different resolution.

Margaret: need at least one interlaced format, to avoid bad surprises later on.

Scope: is working at NAL level good enough ? What information do we lose by doing that ?

Marcus:

- packet info data (not available in bitstream, timestamps...)

- bitstream

- PVS

- SRC

Marcus : should ask at what information is needed by all model so that we do not a test that will be useless because one information was not collected.

Nicolas: when making PVS, should save all the information available (PCAP file...)

Patrick : Need action point to list what's need as input to model to deal with transmission errors.

Margaret: if start with no transmission error test (coding error only), allows to do fast simple test.

Margaret : should start with two small tests, so to check that all modules can handle changes in the 3 variables (resolution, progressive / interlaced, framerate)

Margaret: CDvial website, all members are encouraged to shoot video and upload.

Ugent: Maybe better to start with complete decoder, not NAL parser, because will need more information with time. Reference software.

Marcus: would like to see a vmware image with all tools to produce HRCs

Patrick: on HRCs, coding error will always be useful in the future, but requirements for transmission error may change.

Write the specs for the test, get it approved by the group.

Patrick: Dataset available to the group, with PVS, bitstream, MOS score. So metrics can be tested against it starting from now.

Filippo: One fast way to test the metric is to test them against full ref models.

**Action points:**

- SRC collection, Margaret, Lucjan
  - subjective methodology studies, Ulrich, Filippo, Lucjan, Roland,
  - H264 NAL parser, Ugent, Ircyn
  - metric framework (inputs...) to handle transmission errors, Marie-Neige, Marcus, Ugent
  - software issue, Roland, Marcus, Margaret, Ugent, Lucjan, Kjell
  - first core experiment: available database metric
  - PVS generation =>different encoder (collect PVS generators, report on them), Marcus, Ugent
  - decoders =>SoA and collection, Patrick
  - next call for evidence, Margaret (test specs)
- 

END VQEG Notes