

Final VQEG meeting minutes

Paris May 7-11, 2007

Including Agenda, Participants List, and Minutes from each day's sessions.

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

Agenda

Video Quality Experts Group

May 7-11, 2007 - Meeting

Paris, France

www.vqeg.org

(Note: The ITU-T JRG-MMQA met coincident with the Multimedia and Hybrid Sessions of the VQEG Meeting.)

Draft Agenda

Monday, May 7

- | | |
|------------|--|
| 8:00 | Start / Designate Note Taker (Webster/Speranza)
Introductions
Meeting Logistics
Updates (Maximum 15 minutes - each Group)
Independent Lab Group (ILG)(Brunnstrom/Cermak)
RRNR-TV (Bourret/ Lee)
Multimedia (Brunnstrom/Hands)
HDTV (Pinson)
Tools and Subjective Labs Setup Group (Le Callet)
POC for Source and HRC Sequence collection (Lee)
Hybrid – Perceptual/Bitstream (Juric/Lee)
Calibration Verification (Webster/Lee) |
| 10:00 | Begin MM/JRG-MMQA |
| 11:30-1:00 | Lunch |
| 1:00 | MM and JRG-MMQA
Review Scene Pool
NTIA
Psytechnics
Test Designs
Software status update
Test Schedule
Nortel Contribution |
| 5:00 | End of Day |

Tuesday, May 8

8:00 Start / Designate Note Taker
Review Monday's Decisions (Webster/Speranza)
MM Continues

5:00 End of Day

Wednesday, May 9

8:00 Start / Designate Note Taker
Review Tuesday's Decisions (Webster/ Speranza)
Review MM Decisions

11:30 Lunch

1:00 Hybrid/Bitstream Project Discussion
Verify Current Co-Chairs
Volunteer for Test Plan Editor
Discuss

4:00 Calibration Discussion

5:00 End of Day

Thursday, May 10

8:00 Start / Designate Note Taker
Review Wednesday's Decisions (Webster/ Speranza)
RRNR-TV
HDTV

11:30-1:00 Lunch

1:00 – 5:00 HDTV – Continued

5:00 End of Day

Friday May 11

8:00 Start / Designate Note Taker
Review Thursday's Decisions (Webster/ Speranza)
Final Review of HDTV decisions (if needed)

10:00 Review and Documentation of VQEG Decisions
Review action items

11:00	Other Business Next Meeting
11:30-1:00	Lunch
1:00-3:00	Write Liaison statements to ITU etc.
3:00	Close VQEG meeting

Video Quality Experts Group
May 7 – 11, 2007
Paris, France

Final Participants List

Aladine	Chetouani	L2Ti Univ. Paris
Azeddine	Beghdari	L2Ti Univ. Paris
Barkowsky	Marcus	Opticom
Barriac	Vincent	France Télécom
		FirstStar
Bottoms	John	Networks
Bourret	Alex	British Telecom
Brunnström	Kjell	Acreo
Carnec	Mathieu	Qualidio
Cermak	Gregory	Verizon
Colin	Cyrille	Nortel
Cotanis (Audio)	Irina	Ericsson
Corriveau	Phil	INTEL
Dhondt	Yves	Ghent University
Ford	Carolyn	NTIA/ITS
Garcia	Marie-Neige	Deutsche Telekom
Hands	David	British Telecom
Huynh-Thu	Quan	Psytechnics
Juric	Pero	SwissQual/Spirent
		Universite de
Larabi	Chaker	Poitiers - SIC
Le Callet	Patrick	IRCCyN
Lee	Chulhee	Yonsei Univ.
McCarthy (Audio)	Sean	Modulus Video
Moens	Stefaan	Ghent University
		DVC Digitalvideo
Näther	Harald	Computing
Okamoto	Jun	NTT
Pastrana	Ricardo	Orange-FT
Pinson	Margaret	NTIA/ITS
Planche	Jean-Michel	Witbe
Raake	Alexander	Deutsche Telekom
Rolland	Paul	Witbe.net
Schmidmer	Christian	Opticom
Speranza	Filippo	CRC
Staelens	Nicolas	Ghent University
Sugimoto	Osamu	KDDI R&D
Takahashi	Akira	NTT
Thorpe	Leigh	Nortel

Watanabe
Webster

Wyckens
Yamada

Keishiro
Arthur

Emmanuel
Toru

NTT
NTIA/ITS
France Telecom
R&D Orange
NEC

Monday 7 May 2007

(Thanks to Quan Huynh-Thu of Psytechnics for taking notes on Monday.)

Introduction of meeting attendees

Agenda and logistics review.

Updates from the different ad-hoc groups:

- ILG (KB).
 - o MM models submitted by Opticom, Psytechnics, Yonsei, KDDI, Spirent, NTT and Genista. Yonsei have exercised their right to fix bugs in their submitted software. Software modifications are being monitored by CRC.
 - o Models verification by ILG done
 - o ILG has released a MM source pool
 - o Phase of subjective test designs is not complete yet and will be discussed during this meeting
 - o Fee payment: pairs of ILG lab and proponent have been assigned
 - o Comment by ILG that workload is really big with limited resources and ILG find it is vital that future test plans limit the amount of work for ILG.

- RRNR-TV (CL):
 - o Current timeline:
 - Model submission: Nov 7 2007-05-07
 - Subjective testing by ILG: Dec 15
 - Data analysis: Feb 15 2008
 - Final report: Apr 15 2008
 - o Discussion around the RRNR-TV project:
 - CS said RRNR-TV had to finish quickly otherwise VQEG face the issue of lack of ILG resource to run the tests because of the number of VQEG projects running in parallel.
 - FS commented that it was impossible for ILG to handle subjective tests of the different projects at the same time.
 - DH suggested that one route could be to delay RRNR-TV project and combine it with the 'Hybrid-bitstream' project as the results from the same subjective experiments (at least for SDTV) could be used to evaluate both (pixel-based) RRNR-TV models and 'hybrid-bitstream' models.
 - DH mentioned that RRNR-TV is still held back by the issue of content, i.e. how much of the MM content could be used in RRNR-TV since some MM proponents who are also RRNR-TV proponents would get an advantage of content exposure.
 - Potential proponents: NEC, BT, NTIA and Yonsei (and possibly Genista and TDF).

- MM (KB, DH):
 - o MM models have been submitted
 - o A pool of MM scenes has been distributed and will be reviewed

- Issue about the different email reflectors needs to be discussed
- Since model submission, different contributions have been submitted:
 - Test design by CRC/Nortel
 - NTIA/ILG have submitted a list of proponents tests
 - Psytechnics has submitted a contribution covering an issue in the current test design
 - Opticom and Psytechnics have both submitted contributions covering issues with the current selected scenes in the pool
 - Opticom and Psytechnics have both submitted contributions covering issues and feedback on the subjective test software
- HDTV (MP):
 - No progress since the last meeting
 - Current co-chairs are not active anymore and there will be discussion to replace the co-chairs
 - FS asked about the possibility to merge RRNR-TV project with HDTV project. DH commented on the issue of the different subjective testing methodology used in the 2 projects.
- Tools and subjective lab set-up (PLC):
 - Issues about the ACREO software have been found and reported
 - Issues about the creation of PVSs from the 12-sec sources. Tools are needed to create the PVSs.
 - HDTV: Intel, NTIA and IrrCyN have equipment to run HDTV subjective tests
 - PLC commented on the possibility of merging RRNR-TV and HDTV test plans: HDTV will focus mostly compression and post-processing errors, whilst RRNR-TV should focus more on transmission errors.
- SRCs (CL):
 - KBS agreement ends on 31 December 2008
- Hybrid/bitstream (CL/PJ):
 - CL presented some slides to kick-off discussion
- Calibration verification (CL/AW):
 - AW gave an overview of on-going activities within ITU-T SG9 and ITU-R WP6Q on standardization of video registration techniques.

MM discussion

Topic1: MM scene pool

Updated list of scenes in the pool: 25 April 2006. Content made available on that day via the protected ftp server.

QH presented the Psytechnics' contribution covering issues found in the some of the proposed scenes in the pool. MB presented the Opticom's contribution on the same subject.

The group discussed the issue of having a given SRC included in both the common set and the non-common set of SRCs.

Organizations supporting the possibility to have more than one version of the same SRCs in the same test: ACREO, Spirent, BT, NTIA. INTEL, NORTEL, FT, Opticom, KDDI, Yonsei, FirstStar.

No organization supported to have more than two versions of the same SRC in the same test

Decision reached:

1. The same SRC content can be included in both a common set and non-common set of SRCs within the same test.
2. The name of a SRC has to be unique.
3. No more than 2 versions of the same source can be included in a single test.

MP presented examples of common PVSs generated from the sources in the scene pool.

Discussion around the lowest quality of common PVSs.

Discussion around the common set of PVSs: should each common SRC be run through the same HRC or can different common SRCs be run through different HRCs to produce the common set of PVSs?

Decision reached: for the common set of PVSs, these can be created by passing the different SRCs from the common set through different HRCs.

Discussion about balance between 25fps and 30fps tests:

- MP mentioned that (1) not enough 25fps sources are available to run more 25fps tests and (2) some of the sets of sources do not cover 6 different content categories, especially in the 25fps sets.
- No organization objected in going forward and using these sets although they do not include 6 content categories.

Issues with the current scenes in the pool:

For the scenes in which deinterlacing problems were identified, VQEG did not review them during the meeting but a small group of organizations, referenced subsequently as Reprocessing Committee (NTT, BT, Psytechnics, Opticom, Yonsei, NTT, led by Marcus) will try to reprocess them to suppress the artifacts.

VQEG reviewed the following scenes:

- CommonSetCif:
 - NTIA_SRC_WashdcStill_cif.avi: keep
 - CU_SRC_presents3_cif.avi: keep but also give to Reprocessing Committee
 - CU_SRC_bbfoul_cif.avi: keep but also give to Reprocessing Committee
- CommonSetVga:
 - KBS_SRC_newsG_vga.avi: keep but also give to Reprocessing Committee
- QCIF_D:
 - OPT_SRC_021_qcif.avi: give to Reprocessing Committee
- QCIF_I:
 - IRCCyN_Gob2_qcif.avi: give to Reprocessing Committee
- CIF_G:
 - NTIA_SRC_fcnstop25fps_cif.avi: keep

- CIF_H:
 - OPT_SRC_020_cif.avi: keep
- CIF_J:
 - KBS_SRC_leeparkA_cif.avi: give to Reprocessing Committee
- CIF_N:
 - KDDI_SRC_3D09_cif.avi: give to Reprocessing Committee
- Cif_O:
 - NTIA_SRC_bpit1_cif.avi: shift sequence by 1.75sec so viewer sees a portion of the content without the net.
- CIF_X:
 - SMPTE:SRC_birches1_cif.avi: keep
- VGA_C:
 - T_W_05p_v.avi: keep
 - OPT_SRC_004_vga.avi: keep
 - OPT_SRC_008_vga.avi: replace
- VGA_E:
 - OPT_SRC_009_vga.avi: give to Reprocessing Committee
- VGA_F:
 - OPT_SRC_006_vga.avi: replace
 - OPT_SRC_001p_vga.avi: keep
 - ITU_SRC_CalMobB625_vga.avi: give to Reprocessing Committee
 - NTIA_SRC_ftballslow_vga.avi: keep
 - T_W_10_v.avi: keep
- VGA_H:
 - OPT_SRC_020_vga.avi: keep
 - OPT_SRC_008_vga.avi: replace
- VGA_K:
 - KBS_SRC_winterA_vga.avi: keep
- VGA_L:
 - CRC_SRC_carrousel_vga.avi: keep
 - KDDI_SRC_SD19_vga.avi: keep
- VGA_M:
 - KBS_SRC_soccerB_vga.avi: give to Reprocessing Committee
 - KDDI_SRC_SD16_vga.avi: keep
- VGA_N:
 - KDDI_SRC_3D09_vga.avi: give to Reprocessing Committee
 - SMPTE_SRC_birches2_vga.avi: keep
- VGA_O:
 - CRC_SRC_FlamingoHilton_vga.avi: keep
 - KBS_SRC_newsA_vga.avi: give to Reprocessing Committee
 - KBS_SRC_newsBp_vga.avi: give to Reprocessing Committee
- VGA_P:
 - KDDI_SRC_3D02_vga.avi: give to Reprocessing Committee
 - KDDI_SRC_SD14_vga.avi: give to Reprocessing Committee
- VGA_Q:
 - KBS_SRC_morningBp_vga.avi: replace
 - NTIA_SRC_stadsc_vga.avi: keep
 - KBS_soccerD_vga.avi: give to Reprocessing Committee
 - Yonsei_SRC_zooA_vga.avi: keep
- VGA_R:
 - CRC_SRC_CaesarsPalace_vga.avi: keep

- KBS_SRC_soccerC_vga.avi : give to Reprocessing Committee
- ITU_SRC_MobileCalendar_vga.avi: give to Reprocessing Committee
- VGA_S:
 - KBS_SRC_jeeparkC_vga.avi: give to Reprocessing Committee

Tuesday 8 May 2007

(Thanks to Greg Cermak of Verizon for taking notes on Tuesday.)

TUESDAY

Review of yesterday's notes.

MP regarding test scenes discussed/replaced yesterday. proposing 4 new scenes as replacements for those rejected yesterday.

KBS_SRC_mudbank_vga to be sent to committee for review.

SVT_SRC_closeuplegs2_vga

SVT_SRC_FirstGirls2_vga

SVT_SRC_OldTownCrossPP_vga

Some discussion by Marcus and Quan about details of obtaining and fixing the scenes to be reprocessed by the de-interlacing/reprocessing committee. DECISION: Marcus will try to get the new source content by week of May 21 (most content is already on hand). Then one week for reprocessing (June 2), one week for internal committee review (June 9), one week for VQEG review (June 16). New content will be on FTP site under a folder called "Reprocessing."

Presentation by Leigh Thorpe of Norte regarding video sample production (written by Tim Rahrer):

Not even going to try to describe the detail of the lab setup. Setup is for inserting packet loss using empirical packet loss burst distributions. CIF and QCIF content used to represent current mobile video operating conditions. Outcome depends a lot on the particular player, e.g., loss concealment mechanisms, stability in the face of packet loss. Some packet loss results only in frame loss. Questions raised about the SQQTClient Capture Tool (from SwissQual). Displays the QCIF production matrix of parameter settings to produce 136 sequences plus 16 further sequences. There was a similar CIF production matrix not displayed. In questions from audience: How much diversity in frame rate does a model have to cope with? Answer seems to be that a model should be able to handle a 'reasonable' amount of frame rate diversity. Marcus: Proponents need to supply very much information about HRCs, especially about error-handling. This presentation is available on FTP site.

Later: Quan and Margaret show examples of the Nortel PVSs. The bursty 1% packet loss showed up mainly as occasional distorted blocks.

Test Design:

Psytechnics (Quan). Currently we have 8 SRCs X 17 HRCs + 24 common PVSs. Notes that 24 PVSs are created from 6 common SRCs by 4 HRCs. Including the hidden reference as a 5th HRC

would give a total of 166 PVSs (136 + 24 + 6). Proposes that this be made explicit in the Test Plan. Filippo says the hidden reference is implicit in the Test Plan (by having 3 HRCs); Quan's proposal means adding an extra common set HRC. The new proposal is to have 6 SRCs and 5 HRCs as the common set, one of which is the hidden reference. New common set would be 30 PVSs in common set. Voted (by over 2/3 majority) to amend the MM Test Plan.

Nortel-CRC test plan (Filippo). Only Nortel-CRC and Psytechnics have their designs in; all others are late. Filippo strongly encourages all others to submit their test designs. Also, we would like to avoid duplication in designs. Displays the design matrix. Shows several levels of quality as well as quality levels that are of intrinsic interest to their industry. DH encourages all organizations to supply test designs. Opticom has their test design. DH suggests a target date of May 18 for getting test designs in to Arthur, Margaret and Filippo: Spirent, KDDI, NTT, Yonsei, Genista. Nantes and NTIA/Verizon will produce their designs after the proponent designs are in. Opticom suggests that design-makers include an eyeball-estimate of MOS for the various PVSs.

Software for running tests:

Psytechnics. Regarding Acreo software. Worked on CIF and QCIF, but problems with VGA. (1) Software crashed during VGA. (2) Some distortions were introduced. Supposedly 256 meg on the video card (?) are required. Also, suggests that next PVS be presented once a vote is entered, rather than requiring a click to start the next trial. Suggests that controlled randomization be added to software. Also suggests having black borders around video to be added to software.

Kjell notes that current version doesn't check file names; that VGA does require 256 meg on video card; the extra click; the border; all these are to be corrected. Filippo recommends external randomization software rather than randomization within the Acreo software. DH suggests presenting the randomized common set AS A BLOCK after the practice trials but before the non-common source. LT (Nortel) suggests randomizing across the whole set of common + non-common sets. Suggests externally-produced randomization. DECISION: The group opinion is that there should be a full randomization across both common and noncommon sets of PVSs.

Kjell says he will add a parameter to disable randomization if the user wants, and users can supply external randomization. Margaret says the current player does not handle both 25 and 30 fps; Kjell says it does handle both frame rates. Kjell will mail around the name/location of that player. Having a reasonable file-naming convention should disambiguate common-set and noncommon-set sequences.

Composition of MM Reflector:

Are we happy with the MM Proponent Reflector list? Christian goes through Proponent list. Toyama is a potential MM Proponent, but is not currently a proponent. There also is an MM Test Reflector which is more general. An issue is that general issues are getting onto MMProp, so the distinction between MMProp and MMTTest is becoming blurred. So, do we need MMProp? Proposal: Keep parties on MMProp who have signed the confidentiality agreement or who are in the process of signing. Group seems to agree, although there has been no vote. TDF deleted. Toyama deleted. Lucent deleted. Intel retained (Phil). Nortel retained (Leigh added). Genista list adjusted. BT deleted. Ericsson retained. Other not named were retained if they were on the list.

MM Test Plan & Schedule Version 1.16:

Target date for test designs due now is May 18. Target date for review of Test Plan to be completed is proposed to be June 16, the same day the final source sequences are due. What to do about groups that miss the due date for the test design? Proposal: Those proponents who do not meet the June 1 due date for test design will not be allowed to do the MM subjective test. Margaret notes that the defaulters may still get to see their model validated, but will not get to have access to the processed sequences. So voted 12-1. Addendum: Also noted that if they want their model validated they will have to pay an ILG lab to design and run a subjective test; this will also allow them access to the processed sequences and the subjective data for the MM test. Arthur is word-smithing this statement for posting to all the VQEG reflectors.

Date for sending out fee payment matrix is May 11. Filippo and Greg conferred on payment matrix; Filippo will send it out.

Date for ILG sending invoices to proponents is 1 July.

Other points reviewed but not changed.

Date for signing NDAs reviewed, left as-is.

Point 22. See notes above.

Point 23. Filippo says ILG will not take responsibility for calibration of PVSs; the proponents are responsible for PVS calibration to Test Plan specs. Does someone have calibration software to share? NTIA offers to provide registration software. Discussion about properties of this software. It provides an overall pass-fail result, but is not infallible. Opticom and Yonsei will also make their software available to the ILG. Logistics of shipping all PVSs around discussed. Proposed date for generating PVSs is 1 Sept. 2007.

Point 24. Method of exchanging video data: Filippo notes that since each proponent needs 50% of content from elsewhere, proponents are essentially paired. These pairs of proponents would then be checking each sequence twice (once each) so that the checking process might be handled without video data having to be distributed out to everyone. Therefore, the date for generating PVSs is Sept. 1 and the calibration checks will be done by the time of the next VQEG meeting in Sept. It is noted that organizations generating HRCs should generate a few backup HRCs since some PVSs may fail the calibration check.

Point 25. Date for contacting ILG about problem experiments will be the time of the next VQEG meeting in Sept.

Point. 26. Date also is time of VQEG Sept. meeting.

Point 27. ILG performs validity checks of models. Point 27 deleted because it's redundant with point 29. Points now renumbered.

New Point 27. Proponents run their models and submit their objective data to ILG by Oct. 15.

New Point 28. Results of tests submitted to ILG. Subjective tests finished by 30 Nov.

Point 29. Verification of submitted models by ILG by 1 Nov.

Point 30. ILG distributes subjective and objective data to proponents by 15 Dec.

Point 31. Optional mapping coefficients by proponents submitted to ILG by 15 Jan. 2008.

Point 32. Statistical analysis by 31 Jan. 2008.

Point 33. Draft final report 28 Feb. 2008.

Point 34. Approval of final report 31 March 2008.

Wednesday 9 May 2007

(Thanks to Carolyn Ford of NTIA/ITS for taking minutes on Wednesday.)

Finished MM discussion

- + Margaret presented current subjective test list
- + After test designs have been submitted, an audio call will be arranged to examine the test designs to set the specifics in terms of codecs, etc. and to agree on proponent pairings.
- + Discussion of fee matrix. Currency and schedule of payment to be worked out between the organizations involved.
- + Plan for next round of MM will include audio, so members are asked to start collecting applicable material. VQEG will skip the audio-only test and go directly to audio/video, using audio results from the ITU.
- + Proposal to let FT and IRCCyN run tests with both ACR and SAMVIQ and compare methods (but only use ACR for MM results). Same subjects will be used for both tests, but each half will do the tests in different orders. Only the ACR-first subjects' results will be used for MM. Same test sequences will be used. Info on SAMVIQ can be found at: EBU website. No violent objections.
- + Upcoming dates:
 - o June 16 – source video corrected
 - o June 1 – test designs due
 - o June 16 - Final review of test plan (audio call)
 - o Sept 1 – PVS generation
 - o Next meeting – PVS distribution

Housekeeping issues

- + Phil Corriveau takes over for Viviak as co-chair of HD
- + New co-chairs for tools/subjective labs setup: Ricardo Pastrana from FT and John Bottoms from FirstStar.

Discussion about combining RRNR, HDTV, Hybrid tests

No decisions made. Feeling is that Hybrid test has not been developed enough to make a decision about combining it.

Interest in SD RRNR TV: NEC, Acreo, KDDI, Opticom, Psytechnics, Ghent, NTT, NTIA, BT, Verizon, Intel, Spirent, Yonsei, FT, QualiDeo, (TDF? Genista?) [proponents, ILG]

Interest in HDTV: IRCCyN, FT, DT, Nortel, NEC, KDDI, NTT, Witbe, QualiDeo, Yonsei, Ghent University, Verizon, BT, Psytechnics, Opticom, NTIA, Intel, Acreo

Hybrid perceptual bitstream: IRCCyN, FT, DT, NEC, KDDI, NTT, Witbe, QualiDeo, Yonsei, Spirent, Ghent University, Verizon, BT, Psytechnics, Opticom, NTIA, Intel, Acreo

Hybrid bit-stream model discussion

NTT made a presentation proposing one method

BT made a presentation regarding the scope possibilities

Yonsei made a presentation regarding the definition, leading to a PHILosophical discussion of quality.

{presentations should be posted to the VQEG ftp site}

Long discussion over scope. More or fewer models?

What kind of data can be used as input? Proposal:

h.264-rtp

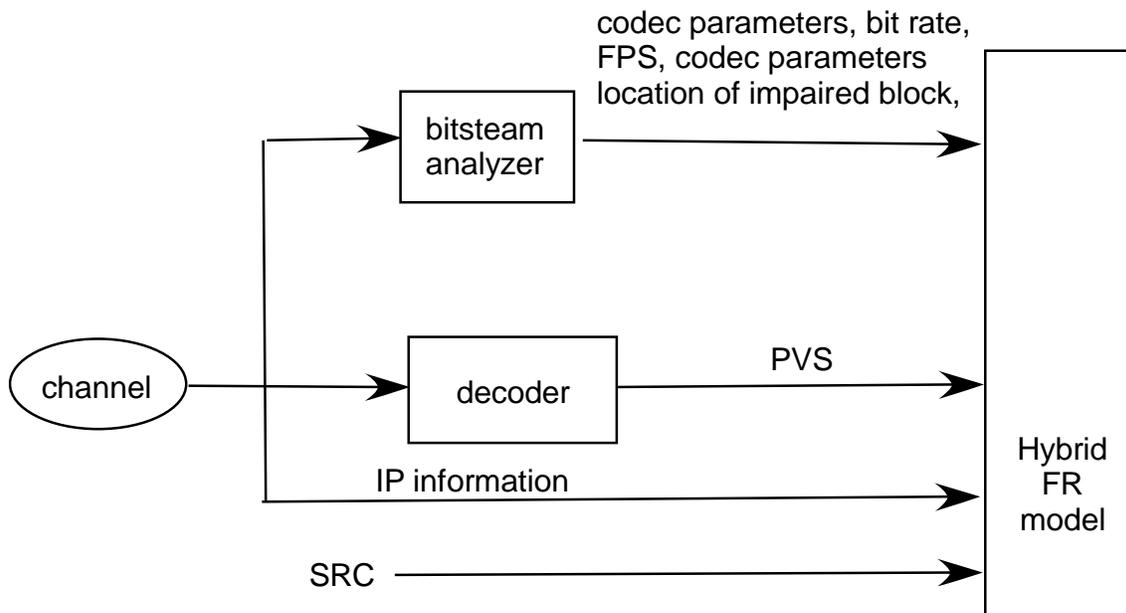
mpeg2-ts

vc1-rtp

vote that TS description be available as freely available software at no cost

Standard set of parameters to be used, or free-for-all?

Latest tentatively accepted proposal:



- ✚ Subjective test: ILG does as little as possible, only what is necessary to ensure fairness for all proponents.
- ✚ Method: ACR w/ HRR (tentatively) for both CIF/VGA and SD/HD
- ✚ Conditions and sources from MM will be recycled as much as possible. It is requested that more source be obtained that doesn't expire.
- ✚ Software tools are needed to capture bitstream data.
- ✚ Further discussion will be had via email and voice conferences.
- ✚ David Hands (BT) and Nicolas Staelens (Ghent) were named Co-editors of the Hybrid Testplan.

Thursday 10 May 2007

(Thanks to Greg Cermak (Verizon) for taking minutes on Thursday.)

Review of yesterday's notes – the decisions:

Some tentative agreements, but no decisions. Note that in the discussions of the Hybrid project yesterday, a Bitstream Analyzer was assumed, but a real, standardized, consensus tool does not exist.

Vincent Barriac (FT) and Akira Takahashi (NTT) liaison from ITU SG12 re Q14 and Q13: Q14 is doing PNAMS which is quality models for multimedia including VoIP that do not use the actual signal stream, but rather parameters derived from the signal stream (e.g., coding type, bit rate, packet loss). Q13 is a network planning model using similar parameters.

Are there points of overlap, mutual interest between SG12 and VQEG? DH: P.564 work may be of use to VQEG in hybrid work. DH and AW are inviting SG12 help in the hybrid project. Apparently SG12 has more expertise in the objective parameters of a media stream and VQEG has more expertise in subjective testing. The SG12 model does not include payload information, only bitstream information. It may be that the future Hybrid Model might look like PNAMS plus information from the content payload (“perceptual information” or PVS).

Another liaison from SG12 announcing that G.1070 (G.O MV) has been approved. It's on the VQEG ftp site.

Looking for tools for dealing with transport streams or tools for extracting parameters from transport streams. DOES VERIZON HAVE EXAMPLES OF MPEG 2 TRANSPORT STREAMS FOR VQEG USE?

What are the next steps for the hybrid project? Nicolas and David need to start a Test Plan. Arthur will set up a Reflector for the Hybrid group; everyone present wants to be on the reflector (hybrid@its.bldrdoc.gov).

Agreements/Decisions regarding the Hybrid Test:

- * Video formats already agreed: (CIF, VGA), (SD, HD)
- * Codecs, etc. provisionally agreed: MPEG2 TS, H.264 RTP, VC1 RTP, MP4 multiplexing; details to be added.
- * Model types agreed: NR(with bitstream info), parametric bitstream (w/out PVS)
Note: FR & RR inclusion is still to be decided.
- * Subjective testing provisional agreement: ACR-HRR for CIF and VGA; ACR-HRR for SD and HD.
- * Test condition and source pool. Agreed that test conditions of MM etc... (missed the rest)

For more details on the decisions see updated slides on ftp site (070509 hybrid models._Decisions.ppt).

RR/NR TV

Margaret Pinson & Alex Bourret presenting slides of Chulhee. New proposed schedule to take account of ILG load. Scene and HRC selection by ILG by Nov. 30, 2007; also model submission

Nov. 30. Creation of PVSs by Dec. 31 (done by proponents under supervision of ILG). Subjective testing by ILGs Jan. 31, 2008; 525 – ILG or NTIA/Yonsei; 625 – T&W. Data analysis Feb. 15. Report Feb. 29, 2008. DH does not want to see subjective tests run only by proponents.

Frequent scene cuts must be allowed in SRCs. At least 1/3 of viewers must be run by ILG. Agreed

Proponents must donate new SRC video by July 31. Also, provide details of HRC systems they can produce. Also sign all NDAs. If no new material can be found and must be purchased, ILG will find and purchase video source material if it's not expensive; ILG would then own the source material and would distribute to proponents if feasible. Agreed

ILG have limited duties. Agreed

Models submitted by Nov. 30, 2007. Preparation of test tapes, instructions, other test materials within 2-3 months after models submitted. Agreed

Each test (525, 625) to be run by multiple labs for cross-lab verification (i.e., ILG runs at least 1/3 of subjects for each test). The preference is for ILG labs to run the subjective tests, but if that isn't possible, then up to 2/3 of the subjects in each test can be run by proponent labs, the rest by ILG. Agreed

Tighten calibration limits (more details). Margaret would like to specify that delay could be more than +/- 2 frames in the case of transmission errors only. Tighten up luminance gain to +/- 3% and offset to 10. All PVSs will be run through calibration software such as that which is being considered for ITU standardization as J.cal TD421 from Oct. 2006 SG9 meeting. Calibration results from 2 or more algorithms will be combined (averaged). All models will accept calibration corrections as optional input parameters as needed. Agreed

Change (increase) RR bit rates: New bit rates are 15 kbits, 80 kbits, 256 kbits. Agreed

Subjective test method: Get rid of SSCQE. Agreed

Subjective test method: ACR with hidden reference removal. Agreed.

Scene length (new): 8 seconds. Transmission errors that impact delay cannot occur in last 1-sec or first 1-sec of test scene. SRC used for PVS creation will have extra 2-sec at beginning and end for editing purposes. Agreed.

One 525-line and one 625-line experiment will be conducted, each containing 160 video sequences. These will be ILG tests. Agreed.

Percent of HRCs that are based on transmission errors remains the same as previous version of test plan (25%), but the absolute number may be different. Provisionally agreed.

Numbers of SRCs and HRCs may be changed to account for change from SSCQE to ACR-HRR. 8-12 SRCs and a number of HRCs to be determined by the test design. No restrictions on test design matrix. Agreed.

Data analysis from MM Test Plan will be used. Agreed.

RRNR proponents were polled about proposed changes above. Proponent TDF is not present.

RRNR editorial committee (A. Bourret, M. Pinson) will finish editing the RRNR Test Plan by May 25 and distribute it. VQEG will have 2 weeks to review edits and propose corrections. If needed, an audio call will take place soon after to finalize the revised RRNR Test Plan.

See Annex I for more details on the RRNR Agreements.

HDTV SESSION

HDTV Test Plan: Margaret proposes simplifying the test plan to be similar to MM and (now) RRNR in order to speed up the project. Specific proposals:

1. Limit HDTV to coding impairments (no transmission error), limited to what is on hand for now.

Discussion: Market wants tools for transmission errors, not coding errors. Others say that HD coding is indeed a big issue among current manufacturers. Patrick points out that if the HD test plan includes many conditions, then we will have to transport those many uncompressed PVSS around.

Organizations that could create transmission errors for HDTV = Opticom, NTT, Nortel. Frame rate resolution possible: NTT = 1080i, 30 fps; Opticom = Any, any; Nortel = have to check.

Comfortable with proponents creating all transmission errors = 5; not comfortable = 2. Who demands transmission errors in first HD test? 3 organizations. Who would be ok with HD that only has compression errors – as a first step? 7 organizations.

Vote: One-phase HD test covering compression + transmission errors : Acreo, Opticom, NTT. Two-phase with coding errors in the first phase: NTIA, BT, Qalidio, Intel, Nantes, FT, Nortel, NEC, KDDI.

One or two test plans? Phase 1 Test Plan will be done first. Agreed.

2. Use 720p and 1080i for source that can be acquired quickly. What about 1080p? Opinion vote re interest: 720p = 8; 1080i = 11; 1080p = 6. In favor of using only 1 or 2 resolutions in Phase 1 = KDDI, NTT, NEC, FT, Acreo, Nantes, Intel, Qualidio, BT, Ghent, NTIA. In favor of not restricting resolution = Opticom.

Who can create coding impairments for 720p? = 6; 1080i? = 8; 1080p? = 1-3.

Which pair of resolutions? Vote: Drop 1080p? = KDDI, NTT, NEC, Acreo, Nante, Qualidio, BT, NTIA. Opposed to drop 1080p? Intel.

Decision: Drop 1080p
Test will include 1080i and 720p

Who can provide source at these resolutions? NTIA. For 25 and 50 fps source? FT. Acceptable to purchase source if it were not too expensive: KDDI, NTT, Opticom, Nortel, NTIA. Not acceptable: None.

3. Adopt data analysis section from MM Test Plan. Provisionally agreed.

4. DSCQS or ACR-HRR? D SAMVIQ being recommended & discussed. No decision.

5. Proponents agree to majority of work, as in RRNR Test Plan? Any support? KDDI, NTT, Opticom, Nantes, NTIA, Qualidio. Not able to help? None.

Editors for HD Test Plan: Leigh Thorpe (Nortel) and Greg Cermak (Verizon).

See ANNEX II for further details on the decisions for the HDTV Test.

ANNEX I (Normative) **RRNR-TV Agreements**
Original source document from C. Lee and M. Pinson

- Need to identify ILG lab with time & resources in approximately January, 2008.
 - Need scene cuts to be allowed frequently in SRC – loosen this requirement
 - Need minimum of 1/3 viewers in each test to be run by ILG
1. Proponents have until July 31, 2007, to:
 - a. donate new SRC video
 - b. tell ILG & other proponents what video systems (HRCs) they can produce, with as much detail as possible (e.g., brand, bit-rates, ways of creating transmission error).
 - c. Sign all content NDAs (e.g., KBS, KDDI, Opticom).
 2. Only mandatory tasks for ILG are the following. All are due when models are submitted (November, 2007).
 - a. Choose (identify) SRC from those provided
 - b. Specify HRCs for tests
 - c. Supply secret SRC if possible
 - d. Specify secret HRCs if possible
 - e. Verify data analysis if resources permit
 3. If ILG cannot provide secret SRC, then the ILG will identify SRC material that can be purchased by each proponents for a small fee. Such SRC will be identified to proponents and purchased by them after model submission. Alternatively, ILG may purchase directly such SRC, if the fee is small enough.
 4. Models submitted November, 2007. Proponents receive instructions from ILG on subjective test design (which SRC, which HRC). Proponents at this point create HRCs, create viewing tapes, request instruction from ILG if test(s) appear unbalanced, distribute video as needed, and run viewers. All to occur within 2-3 months after model submission. All proponents assist in this effort.
 5. Preferably, ILG will run all viewers through all RRNR-TV subjective tests. If ILG do not have the resources available for this, proponents may run at most 66% of viewers any single subjective test; and the remaining 34% of viewers from each individual subjective test must be run by the ILG.
 6. Replace the data analysis with that specified in the MM test plan. Inappropriate provisions will be removed (e.g., references to the common set, references to aggregation of multiple data sets). Such changes will be reviewed during the audio call.

Other issues need to change or consider on RRNR-TV test:

1. Tighten calibration limits, making it clear that calibration algorithms and RR bandwidth will be tested separately by Chulhee's ITU question on that issue. This will require few changes, probably only the following:
 - a. Delay limits currently in RRNR-TV test plan are fairly tight and appropriate at a hard & fast limit of +/- 2 Frames. The exception for "Dropped or repeated frames" should be expanded to include "response to transmission errors". This appears to be the intent, but some transmission errors don't just drop or repeat frames.
 - b. Tighten luminance gain & offset calibration limits to:
 - i. Maximum allowable deviation in offset is +/- 10
 - ii. Maximum allowable deviation in gain is +/- 3%
 - c. All PVSs will be run through calibration software such as those being considered for ITU standardization in J.cal (TD421 from the October 2006 meeting of SG-

- 9). Calibration results from 2+ algorithms will be combined, and exact calibration values agreed upon for each PVs.
 - d. All models take calibration corrections for the processed video sequence as optional input parameters, when needed.
2. Change RR bit-rates to: 15 kbits/s; 80 kbits/s; and 256 kbits/s.
3. Change from SSCQE to ACR-HRR with 8-sec SRC (as pre MM), where transmission errors that impact delay cannot occur in the first 1-sec or last 1-sec of any PVS. As with MM, SRC used for PVS creation will include an extra 2-sec at the beginning and end of the sequence.
4. One 525-line and one 625-line experiment will be conducted, each containing 160 video sequences.
5. Limitations on amount of HRCs containing transmission errors based on numbers of HRCs will be removed. The limitation of HRCs containing transmission errors based on percentage of HRCs will be provisional, to be determined during an audio call, after edits have been entered.
6. Number of SRC and HRCs will be modified as follows: 8 to 12 SRC; and a number of HRCs as determined by test design. Further restrictions on test design (e.g., if a full matrix is required) will be removed.
7. A small editorial committee will make changes to the RRNR-TV test plan to implement the agreed upon changes, to include H.264 (as agreed upon in Tokyo), and to reflect technological changes (e.g., viewing tapes may not be needed).
8. Preferably by May 25, 2007, the editorial committee (i.e., small group motivated RRNR-TV proponents & ILG working on the VQEG reflector) will finish edits.
9. The completed document will be distributed to all of VQEG via VQEG reflector and/or VQEG web site. Other VQEG'ers will have two weeks to review edits and propose corrections. All decisions (from Paris minutes) will be agreed upon here during the Paris meeting, not by the editorial committee.
10. Audio call will be conducted soon thereafter, where any further required agreements can be made (e.g., any issue that has been overlooked). The final RRNR-TV test plan will be approved at this or a follow-on audio call.

ON HDTV
Original proposal from C. Lee and M. Pinson

1. DSCQS vs. ACR-HRR vs. Samviq? Interest exists in all three methods. No decision reached.
2. Limit HDTV to coding (i.e., no transmission errors), limited to what is available to ILG and Proponents. Phase 1 will be coding only; phase 2 will be transmission errors.
3. Limit HDTV to image resolutions for which we can get sufficient SRC in the next few months (perhaps 720p & 1080i).
 - a. Interest = 720p 8 - 1080i 11 - 1080p 6
 - b. Ability = 720p 6 - 1080i 8 - 1080p 1 to 3
4. Adopt the data analysis metrics from MM test plan, marked provisional, modified as needed to suit. Christian will examine and as needed correct the data analysis.
5. Use division of labor similar to RRNR-TV above (i.e., minimal work from ILG required; proponents agree to do majority of work). ILG will choose SRC and HRC; proponents will edit material, run HRCs, and (if needed) run some viewers under ILG direction.
6. If VQEG cannot find sufficient SRC, all proponents & participating ILG will purchase the SRC of ILG's choice (e.g., up to \$4K to \$5K). If this is needed, proponents will find and suggest companies that sell HDTV content.

Friday 10 May 2007

(Thanks to Christian Schmidmer, OPTICOM for taking the minutes on Friday.)

Meeting starts at app. 9:10

- **Discussion and acceptance of minutes from Thursday**
- **HD Testplan:**
 - *Decision on co-editors for HD Testplan (Leigh Thorpe, Greg Cermak) was added to yesterdays notes.*
- **HD Continued:**
 - Discussion on what would be the subjective test method for HD.
 - FT supports SAMVIQ, but there are also other proposals which seem to be valid (e.g. DSCQS). All methods seem to be acceptable in terms of the testing methodology.
 - The main concern with all methods is however the duration of the test which according to Filippo increases from ACR-HRR over SAMVIQ to DSCQS.
 - One disadvantage of SAMVIQ is that only one subject at a time can run the test (pointed out on Thursday by Filippo).
 - Marcus reports experience that SAMVIQ is well suited to evaluate small as well as large degradations within one test. The result was still very accurate.

- Quan has also conducted a SAMVIQ test on data similar to the MM testplan requirements. For very similar qualities in one test SAMVIQ is better suited than ACR-HRR. For a wide range of degradations in one test, both methods were reported to be very similar.
- Marcus points out that the SAMVIQ is more appropriate than ACR if the qualities used in the experiment are not balanced over the MOS scale. This is what may happen in the HD test. Here the distribution of scores may form clusters towards the two ends of the scale. Opinions on this vary significantly amongst the experts. Patrick points out that such a distribution will lead to a good correlation of most models since they only have to predict the cluster to which each sample belongs. He therefore proposed on Thursday to split the testplan into two distinct plans for the two quality ranges.
- Information from Arthur: SAMVIQ is not yet standardised by the ITU. He cites from a Japanese ITU contribution comparing DSCQS with SAMVIQ. The correlation between DSCQS and SAMVIQ is reported to be very high. SAMVIQ is faster than DSCQS according to the contribution.
- KDDI reports experience with MM resolutions only. Prefers DSCQS.
- NTT is in favour of DSCQS.
- FT prefers SAMVIQ
- NEC: is in favour of DSCQS.
- FT reports that stability of SAMVIQ is very high.
- Do we need fully factorial designs with SAMVIQ? Yes, principle is similar to MUSHRA tests.
- Chris raises serious concerns regarding the suitability of SAMVIQ for testing models since the subjects might use different scales for each scene (content) presented to them. This highly depends on the selection of anchors shown with each content type. The models will always apply the same scale independent of the content.
- Arthur: We have three good methods on the table. Purpose is to evaluate models. The output of test will go to standardisation bodies. We can be very conservative and use only standardised methods. This would require DSCQS for HD. However, we don't have to be conservative and we could use ACR-HRR or SAMVIQ. Personally he wants to push SAMVIQ in his lab to gather information for the future. Discussion should continue over the HD reflector.
- Surveys by Margaret:
 - How many people could consider changing from DSCQS to x at the next meeting? 6 pro, 4 against.
 - Who is interested in qualities between broadcast and Multimedia? 9 pro, 0 against.
 - How many clips per resolution we would like to include?
 - Limit to app. 50: 1 pro
 - Limit to app. 160: 4 pro
 - App. 1000+: 2 pro
- Other ideas should be send to Margaret and Filippo
- **Arthur reminds us that we should not act completely independent from the ITU and other bodies.**

- 11:00 25 min Coffee break -

- **Other Business**
 - Video Clarity product presentation and short user discussion
 - Discussion if we want to have unsolicited advertising during meetings. There seem to be large support for not having this. Chris volunteers to write a formal policy on commercial advertising at VQEG meetings and reflectors.

- **Next meeting**
 - Provisionally in Ottawa during last week of September 2007-05-11
 - Proposal was mentioned to have more, but shorter meetings.
 - The wish for more teleconferences was expressed by Phil.
 - NTT still thinks about hosting second next meeting. Boulder seems to be possible as an alternative too.

- 11:45 Meeting adjourned, liaison drafting to follow

End of Meeting notes for VQEG's Paris meeting May 7-11, 2007