

Video Quality Experts Group  
 April 24 – 29, 2005  
 Scottsdale Arizona USA.

Participants:

<b>NAME</b>	<b>Organization</b>	<b>Country</b>
Eugen Rodel	SwissQual	Switzerland
Rene Widmer	SwissQual	Switzerland
Osamu Sugimoto	KDDI	Japan
Kjell Brunnstrom	Acreo	Sweden
Greg Cermak	Verizon	USA
Patrik LE CALLET	IRCCyN	France
Irina Cotanis	Ericsson Netqual	USA
Takaaki KURITA	NTT	Japan
Jun OKAMOTO	NTT	Japan
Christian Schmidmer	Opticom	Germany
Margaret Pinson	NTIA/ITS	USA
Stephen Wolf	NTIA/ITS	USA
David Hands	BT	UK
Quan Huynh-Thu	Psytechnics	UK
Chulhee Lee	Yonsei	Korea
Jean-Louis BLIN	France Telecom	France
Ee Ping Ong	I2R	Singapore
Z.M.Parvez Sazzad	Toyama University	Japan
Yuukou Horita	Toyama University	Japan
Alex Bourret	BT	UK
Arthur Webster	NTIA/ITS	USA
Filippo Speranza	CRC	Canada
Carolyn Ford	NTIA/ITS	USA
Zoe Liu	Nokia	USA
Mingjian Yan	Nortel	USA
Philip Corriveau	Intel	USA
Vivaik Balasubrawmanian	Intel	USA
Jorge Caviedes	Intel	USA
Uday Marty	Intel	USA
Mylene Farias	Intel	USA
Tim Rahrer	Nortel	USA
Pierre Bretillon	TDF	France
William Walton	Johns Hopkins/Applied Physics Lab	USA

Decisions of the meeting are clearly marked in the document and **colour coded**. Thank you to all who took minutes.

# Minutes from the Meeting

Monday April 25<sup>th</sup>, 2005

## Opening of the meeting:

PC thanked VB for arranging meeting.

Participants introduced themselves.

PC covered logistics and schedule.

AW introduced VQEG general rules about decision making and note taking.

PC asked for volunteers to take notes.

VQEG chairs would like to reach decisions at the end of each segment.

AW gave an overview of the agenda and agenda was approved

CL is announced to be the new co-chair of RRNR-TV, replacing

Alexander Woerner.

JC would like to run a subjective test during the meeting but still waiting for equipment to arrive.

## Updates of the different ad-hoc groups:

### ILG update by PS:

PS reported no progress/update from the ILG concerning RRNR-TV because of a lack of availability of sequences:

- Sequences from FRTV-II are not usable (copyright issue with Teranex and Universal sequences).
- CRC-Canada does not have 525-sequences, FUB has 625-sequences from Portuguese TV.
- Material compliant with test plan is difficult to find.

PS said some sequences could be purchased if participants were willing to share costs.

AW said Teranex sequences are available.

AW asked list of organizations that would like to acquire the Universal Studios material.

PS announced that Verizon will run subjective tests for RRNR-TV.

PS underlined the fact that best possible scenario is that subjective tests could only be run in 5-6 months.

Current ILG labs: CRC-Canada, FUB and Verizon.

CRC-Canada received material from Intel and Opticom but said these can't be used in the RRNR-TV due to restrictions imposed by the test plan on sequence length and number of scene cuts.

PS asked to change the restrictions imposed by the RRNR-TV test plan on choice of test sequences because finding sequences compliant with these restrictions is very difficult.

**RRNR-TV update by AB:**

AB met with Vittorio Barroncini (FUB) and it seems that FUB is currently producing the HRCs for 625-sequences. AW reminded that the RRNR-TV test plan is completed and current issues cover access to test sequences and production of HRCs. RRNR-TV proponents: Yonsei, BT, NTIA/ITS, Toyama University, KDDI. Latest version of test plan is version 1.7f. MP asked if the latest changes had been inserted in this final version. AB will check this.

**MM update by DH:**

DH reported 2 audio calls (Dec and Feb) + 1 web-meeting that lead to good progress on the MM test plan (current version is 1.5a). the following chapters were updated:

- Data format
- Transmission errors
- Evaluation metrics
- Objective models
- New section on definition

Target for this meeting is to finalize MM test plan, more specifically there's a need to:

- Agree on latest changes in the MM test plan
- Discuss and agree on color space issue
- Discuss playback software, test procedure
- Discuss access to test material

**Subjective lab set-up group by AW (due to absence of Vittorio Barroncini):**

Vittorio Barroncini is currently the only co-chair. AW asked another volunteer to co-chair that group.

PS suggested to merge the 'Tools' group and this group.

**HDTV update by SW:**

SW said that a skeleton test plan exists but that it does not include any ratified decisions.

SW suggested to use the objective models recommended in ITU-T/ITU-R and test them as they are against HDTV data. SW said he would like to reach decisions during this meeting on test methodology and video format so that participants could start collecting HDTV test material.

Model Combination and Tools Group update by SW:

At the end of FRTV-II, there was a proposal to combine the different objective models into a single model. This work was to be carried out by the Model Combination group. This would need a sort of closed collaboration between the involved companies with a NDA in place. NTIA cannot participate into such a closed collaboration. AW suggested that SW should step down from the 'Model Combination' group. JC said he would like VQEG to focus on the technical work and leave the standards bodies to decide if there is a need to combine the different models into a unique model.

**Decision:** remove the 'Model Combination' ad-hoc group and keep only a 'Tools' ad-hoc group.

**Decision:** Merge 'Subjective test lab set-up' group and 'Tools' group into a unique ad-group group. There will be 3 co-chairs: Mylene Farias (Intel), Steve Wolf (NTIA/ITS) and Vittorio Barroncini (FUB).

SW said NTIA is developing Matlab-based tools and proposed that the Tools group could develop new Matlab-based tools and make them available to VQEG. SW is calling for participants to contribute code to this set of available tools. SW asked any interested party to send him an email with a description of their proposed tool and he will set-up a page on the VQEG web site.

MM pre-test update by AW:

AW reported that some organizations did some pre-test as part of investigations on appropriate subjective test procedure for the MM tests. Latest test results were reported at the Seoul meeting but no further testing was reporting since.

Source and HRC Sequence collection update by CL:

CL presented a summary of the test material that some organizations could provide to VQEG.

Current potential source material: FRTV-I, FRTV-II, SwissQual, NTIA, KBS, MBC (CL is trying to get permission from MBC to distribute these to VQEG).

NTIA needs to have the release form signed by the persons appearing in their video sequences before they can make those sequences public. There were discussions about the availability of audio track in the current source material, as VQEG will address multimodal (audio-visual) testing in the future. If audio track is available, content providers are requested to provide information about the audio track.

## **Proposal from UCSB (sponsored by Intel) to pursue a standard consumer video testing database (presented by JC):**

Intel is proposing to build a standard pool of video sequences (similar to sequences used in MPEG), i.e. high-quality content from which processed video sequences could be produced to benchmark algorithms and applications or run subjective tests. The idea is to have a set of standard sequences to test any video processing out of the coding part itself.

## **Presentation of the HDTV test plan by VB and SW:**

- Introduction on HDTV: scene set-up and motivation
- HDTV source scenes and HRCs: content description, copyright issues and HRC characteristics
- HDTV test methodology:
  - o Subjective testing
  - o Sequence processing

During the presentation on the actual test plan, the following points were discussed:

- Frame rates: KB said that European broadcasters might use 25 and 50 so would like those to be added in the list of considered frame rates. **Decision:** Add 25 and 50 fps.
- Bit rates: current proposal is 8-25Mbps. SW would like to consider lower minimum bit rates (2-4 Mbps). **Decision:** bit rate range will be 2-25 Mbps.
- Packet loss: current proposal is 0-25%. JLB said 25% is too high and unrealistic. PC proposed 10%. TR said that 10% could be considered for wireless transmission but that wired transmission should not suffer more than 1% loss. TR emphasized difference between IP packet loss and video packet loss. VB said packet loss considered here was video packet loss. **Decision:** packet loss range 0-10%.
- JC said he would like to add noise in transmission errors. **Decision:** noise will be allowed in HRCs.
- Decoding and processing: TR would like to add effects of error concealment. **Decision:** HRCs with error concealment are allowed.
- Display technologies: should HDTV subjective tests be performed using different types of display technologies (LCD, plasma, projectors, broadcast CRT...) or only on one type of display? In the case of MM test, LCD was chosen because it is recognized to be

representative of the target applications. However, for HDTV, there are several display technologies that could be representative. JLB said that objective models are tested independently of the display effect. SW proposed to use the most transparent display.

Professional-grade broadcast multi-format CRT monitors could be a solution. Current organizations that could run HDTV subjective tests: Intel, CRC-Canada (and FT if they use their own subjective testing method): Intel has D24 Sony multi-format CRT monitor; CRC-Canada has Sony 16:9 21" multi-format CRT. JC said he would give AW some contacts at Phillips and Sharp to investigate the possibility to have these companies to provide some displays.

**Decision:** ILG will use same type of display. **Decision:** displays should preferably not introduce post-processing or down/up-conversion.

- PS underlined that the number of HRCs in HDTV will have a similar (high) order of magnitude than in the MM test plan.
- Subjective testing methodology:
  - o Current proposal is DSIS: PC asked why DSCQS is not used. SW proposed DSIS because it is simpler and less time-consuming (using variant I with only 1 presentation per trial). MP said that although a bias is introduced by DSIS (due to the fact that reference is always shown first) but bias is unidirectional and known. **Decision:** Double-stimulus Impairment Scale (DSIS) Variant 1 will be used.
  - o Viewing distance: current proposal is 4H. JLB said viewing distance should be dependent on the video format. PC said that consumer trends is to buy bigger and bigger TV sets whilst the size of the room does not increase, thus end-user viewing distance tends to decrease and it has certainly changed over the last decade. **Decision:** 1-min of arc will be used as reference for setting viewing distance (cf ITU-R BT.500-11 page 4).
- Sequence processing: 1080p, 1080i and 720p formats. **Decision:** No up-sampling or down-sampling allowed for the source sequences. SW said that we must be careful that cameras don't introduce any conversion.
- MP asked if up/down-sampling could be included in HRCs? VB said that HRCs can include have up/down-sampling (for example down-sampling applied by the encoder) as part of the test condition.
- **Decision:** Proponents are allowed to conduct approved HDTV subjective testing.

VB covered the key challenges in HDTV testing.

- HD content

- Reliable HD playback for subjective testing
- Investments required for HD testing
- Tools for HRC error injection
- Reference display processing algorithms
- Selection of back channel bandwidth for HD

SW asked to discuss about minimum specs of HD cameras that could be used to shoot test sequences. VB said he would like to allow capture of contents with a variety of HD cameras (not only professional/studio-grade cameras).

Intel proposed a 3-step approach:

1. Subjective testing
2. Evaluation of objective models (starting with the FR models for SDTV already recommended)
3. Results analysis

Discussions on testing existing recommended FR models: MP said that some software development might be needed to make the existing software work for HDTV.

A call for proponents will be sent to the VQEG reflector, ITU-T SG9 and ITU-R WP6Q reflectors.

**Decision:** the existing calibration limits from the FRTV-II test plan will be used.

**Decision:** Objective models' interface will be taken from the FRTV-II test plan and modified for HD.

TR (Nortel) took the action to provide information on network and service deployment so that this information can be considered in HRCs selection.

Organizations with HD playback capability: CRC-Canada, NTIA, France Telecom, IRCCyN

IRCCyN would be able to run HDTV subjective tests in the next 3-6 months if material was available today. CRC is unlikely to be able to. PC suggested limiting 1<sup>st</sup> test to one format.

MP made the remark that if HDTV is to move forward, test plan needs to be completed as soon as possible. She suggested a web meeting.

Tuesday April 26<sup>th</sup>, 2005

## RRNR-TV Session

Terenex (now Silicon Optics) have verbally & email agreed to let VQEG use these sequences. However, they want a letter that their president can sign describing the sequences & outlining usage limitations. Then, we will have official permission to use the Terenex sequences. This should happen in a week or two.

Universal has verbally agreed to renew the agreement from FR-TV Phase II. We are waiting on a final list of the people who need that material. Some proponents might not be able to sign this legal form, as was the case for FR-TV Phase II.

**Agreement was reached:** all MM proponents and ILG should be on the list sent to Universal Studios, to open the possibility of these sequences being used in the MM test.

The KBS material was obtained by Chulhee Lee for the MM test, but could likely also be used for the RRNR-TV test. This is about 50 minutes of sequences (sports, music, etc). An agreement must be signed. There is also the possibility of purchasing a one-hour program for around \$500 per organization, which show can then be used for any purpose.

625-line materials, Vittorio is planning to use a D1 converter on the analogue output of a set top box. This currently might not be within the scope of the test. Alternatives are to have others create error conditions.

Restrictions of temporal registration in the RRNR-TV test plan are perhaps ambiguous. A request was made that new text be added to the test plan to clarify this issue.

**Agreement was reached:** to insert the following text into the RRNR-TV test plan: "Thus, SRC and HRC shall be the same length, and only local temporal variations will be allowed. For example, the +/- 2 frame temporal alignment restriction does not apply to repeated frames resulting from transmission errors."

Text was inserted into the RRNR-TV test plan that had been agreed to at the Seoul meeting.

**Agreement was reached:** to replace calendar dates. "Complete" was added to finished items, and items with uncertain dates were modified to have timestamps occurring after prerequisite events (i.e., "baseline" plus days or months). These timestamps will be separated by the same periods of time established in version 1.7f of the RRNR-TV test plan. The revised test plan will be version 1.7g. This version of the RRNR-TV test plan will be reviewed on Friday, April 29, 2005.

## **MM Session**

Current organizations interested in being MM proponents: Psytechnics, BT, I2R, Opticom, NTIA, NTT, SwissQual, Yonsei University, KDDI, Toyama University, & Genista.

Current organizations that might be test labs for MM test are (ILG): FUB, Intel, CRC, Verizon, Acreo, & France Telecom. (Proponents): BT, Psytechnics, NTIA, NTT, Opticom, Yonsei, Toyama, & SwissQual.

Proposal: no confidence interval for Pearson correlation coefficient. This proposal was not agreed.

**Agreement was reached on:** on all text marked as "accepted" in the MM test plan version 1.5b.

**Note:** Opticom volunteers to provide a program to the ILG that is capable of calculating Monotonic Cubic Polynomial fits, if this is required for the MM test analysis (see Section 8.2.1 version 1.5b).

**Agreement has been reached:** to add a square root to equation (2); and in Section 8.3.2 and Section 8.4.2 replace "PEavg" with "rmse" throughout.

**Agreement has been reached:** to change equation (11) use standard error (rather than standard deviation) & fix the related text to match; there was not a 2/3 majority vote to overturn this previous agreement (i.e. change standard error to standard deviation was rejected).

**Agreement was reached:** to examine and fix the various "N" in section 8 to account for the degrees of freedom lost by the logistics fit, where appropriate; taking into account that the type of fit might change (see section 8.2.1).

**Agreement was reached:** for general information on scene selection & other information on designing tests, please see the video quality tutorial on the ITU web site (link to be provided when available).

Wednesday April 27<sup>th</sup>, 2005

MM Ad-Hoc and JRG-MMQA meetings

This morning will be on presentation of documents. Acreo, Psytechnic, Opticom.

Discussion on aggregation and mapping

Chapter 4

Selection of the color space this afternoon

Presentation by Acreo, on Chin rest and subjective test program  
Subjective test similar to BT's was performed, at a distance of 8H.  
Correlation between chin rest / no chin rest are high, in accordance with BT's test. However, the lack of test points in the middle MOS range is likely to increase the correlation artificially.  
Call for new test covering the middle range MOS.

Presentation by NTT, Proposal of rule for choosing ref and distorted video sequences

For proper and fair comparison between models, should :

- deciding the upper limit ratio of known src and hrc
- either opening phase 2 or not using them in the MM test
- disclosing the properties (organisation offering / processing the sequences) of selected sequences

A discussion must be held in the proponents group on these issues (how much material is kept secret...)

**Action** : Proponents will gather to discuss this tomorrow (Thursday) lunchtime.

Presentation by SwissQual, Capturing video problems, comments on players.

Prefers to use RGB colorspace because easier to display during the test.

Presentation by Psytechnics, on colorspace

- models should be treated as black box
- codecs should be treated as black box
- viewer and models should get same data

Presentation by Opticom,

Subjective test following the rules defined earlier by the group.

- Found out that break is necessary every 10 minutes, about every 60 samples.

- Difference between consecutive sequences minimised by choosing random order that minimise the PSNR difference

- doubled the pixels in every direction to make the viewing distance more comfortable.

**Action** : Opticom to make available some of sequences used for the test.

Presentation by Verizon, on agregation of data across different languages, test conditions.

Logistic fit solves the between labs and between language differences.

What would be agregated is not the raw data, but the correlation, rms error and outlier ratio.

Note : desirable to have some common sequences, in term of content and quality level.

**Agreement was reached:** a common subset of test sequences will be used in every experiment, one per format.

No decision is yet taken on the randomisation / presentation order of this subset block.

NTIA, proposal for conducting MM subjective testing

Following the set of proposal insures that proponents did not train their models on the data.

However, the proposal puts significant loads on the ILG.

Issues :

- source material, spitted into known and unknown

  - known to everyone

  - unknown is only available to ILG

  - known to a specific proponent

- creating HRCs, guideline

  - proponent creates HRCs and use in their test

  - proponent creates HRCs and send them to the ILG

- subset of common sequences has to be defined.

The group will return to NTIA's document later in the week.

**action** : create a list of key decisions to be made based on NTIA's document (Filippo, Margaret)

Subjective assessment (Chapter 4 on test plan), Verizon

**Agreement was reached:** to accept the changes made in chapter 4.1.1, with "for the degraded sequence" inserted.

Changes to chapter 4.1.1. and Annex I were discussed.

Modifications to Annex I :

**Agreement was reached:** The modifications to the "notes" section were accepted. A sentence on adaptation to language and culture was added.

Discussion took place on whether explaining to the subject what the target application or telling him "next generation device"/"multimedia devices".

For Annex I the term "next generation meeting" in the greeting section **is not yet agreed.**

**Agreement was reached:** The changes to the "(vision test)" section were accepted

The two last sentence of the (overview of task) were discussed.

**Agreement was reached:** The last two sentences of "(overview of task)" section were removed. The rest of the paragraph was accepted.

**Agreement was reached:** The changes to the "(physical setup)" section were accepted

**Agreement was reached:** The paragraph on "(room & lighting explanation)" was deleted

**Agreement was reached:** "Suggested" is added to the "(presentation timing and ordering)" section. The section was accepted

**Agreement was reached:** The changes to the "(what you do : judging -- what to look for)" section were accepted. The sentence explaining the different image scale was removed.

**Agreement was reached:** In the "(what you do: rating scale)", the numbers in the scale were removed. First sentence is modified to read "When prompted"...

Second paragraph of section (after the scale) is modified to read "please indicate your rating by pushing the appropriate numeric key on the response pad"

**Agreement was reached:** The sentence allowing to see the sequence again is removed.

Several options to allow subject to change their vote were discussed.

**Agreement was reached:** The note at the end of the paragraph was deleted.

**Agreement was reached:** The section "(practice trial)" was accepted

**Agreement was reached:** The section "(questions)" was accepted

**Agreement was reached:** The (consent form) section was replaced by "where applicable".

**Agreement was reached:** The term "video" is used instead of "image" in the annex I

Chapter 4.1.1 is discussed. Change to be done to figure one to include prompt.

The presence of numerical values in the scale is discussed. Numerical scales have never been used in any ITU test.

**Agreement was reached :** the number are kept in the scale  
Phil objects to the decision and will provide the evidence against it.

(Note: This decision was later reversed.)

**Agreement was reached:** all changes in chapter 4.1.1 are accepted.

Chapter 4.1.2:

First 3 paragraph are minor editorial changes, and notes on headphones deleted.

**Agreement was reached :** all the changes in the first paragraph.

**Agreement was reached :** In second paragraph, ("Presently, VQEGMM assumes..."), all proposed changes are accepted  
An editor note is added, specifying that video card has still to be defined.

Across all section, "601" is replaced by "VGA".

**Agreement was reached :** Final paragraph ("we note regarding...") is deleted.

Chapter 4.1.3

First paragraph, most modification are language clean up.

**Agreement was reached :** all modifications in first paragraph are accepted

**Agreement was reached :** "should" is replaced by "shall" in beginning of second paragraph ("The LCD ...")

Editor's note added on selection of a minimal response time for LCD monitors.

Discussion on availability of systems to measure response time, availability of a given display, calibration of monitors.

**Action :** working group to be setup to define what the specifications

have to be for the display. Kjell agrees to chair the new "Display spec / setup group" to determine monitor calibration among other things. Coordination with Tools Group is expected.

Presence of number in the scale is re-discussed.

**Agreement was reached:** Scales are to be presented to the subject without numbers

A sentence is added in the 4.1.1 on that topic.

Chapter 4.1.4:

**Agreement was reached:** In the first paragraph, the note is deleted

**Agreement was reached:** Changes in second section of the paragraph are all agreed.

Chapter 4.1.5:

**Agreement was reached :** Editorial changes. The words "appropriate viewing distance" are left. The note regarding the possible use of chin rest is removed. The section in 4.1.2 related to viewing distance is copied in this paragraph, and the following sentence is added: "At the start of each test session, the chair will be positioned at the same point (nominally in the middle of the viewing distance, ie 8H for QCIF or 7H for CIF)".

Chapter 4.1.6:

**Agreement was reached :** all modifications to 4.1.6 are agreed.

Chapter 4.2:

Data format defined in 4.2.1. Small edits to the original text.

**Agreement was reached :** all edits in the first paragraph are accepted. "Each row in the excel spreadsheet will contain data for one subject" is added.

**Agreement was reached :** Resolution is changed from "601" to "vga".

**Agreement was reached :** Edit in "order" section: "if scenes are ordered randomly" changed with "if unavailable".

**Agreement was reached :** in the "scene" section, "name shall be ten characters or fewer" is added.

**Agreement was reached :** HRC section, a sentence is added : "HRC will be ten characters or fewer"

**Agreement was reached :** Added a reference to Annex II

**Agreement was reached :** In Annex II, in the order column, the sequence is replaced with "1 2 3 4 ...". A note is added, with -9999 replacing missing information.

## Chapter 4.2.2

**Agreement was reached** : The added note is agreed. "in BT Rec. 500-10" is added after the cross reference to section 2.3.1

Discussion on viewing distance for CIF and QCIF.

Presentation by Acreo, on experiment video player  
Not yet tested for VGA. Does not currently read AVI. Uses directdraw mode which takes over the whole display and controls the display refresh rate. Swissqual and Opticom have routines to read uncompressed avi files.

Discussion on possibilities of fully automatised setup for subjective testing.

Two approaches : force everyone into using the same program, or allow each lab to decide for each program block (ie : display, score selection...).

NTIA presentation on testing different player. Conclusion of the study is that wm9 was the best, and that the group has to define a system.

specification for minimum requirement for pc ?

**Agreement was reached** : The group agrees to have a minimum specification for PC to be used.

**Agreement was reached** : The group agrees to specify a single class of graphic card.

Choice between Yuv / RGB color space was discussed.

Advantage of Yuv : internal of codecs, takes 2/3 of the space on disk, bandwidth required is lower.

**Agreement was reach** : to use the Yuv color space.

Different type Yuv format was discussed : uYvY, YuY2.

**Agreement was reach** : to use the uYvY format.

Section 6.6 contains the format definition. The full definition of AVI / Yuv has to be put there.

**Action** : Vivaik and Margaret are going to define the file format for the avi wrapped Yuv.

**Agreement was reach** : to keep the avi format.

**Action** : Yonsei has agreed to create the test vectors.

**Agreement was reached** : All will use the same conversion matrix between RGB / Yuv. Will be discussed and agreed to on the reflector. All details will be discussed and decided upon on the reflector. Decision needed by end of May.

Thursday April 28<sup>th</sup>, 2005

Review of yesterday's Minutes with last minute additions to the text for clarification.

It was **decided** that the Tools group will provide/adapt the tools to be used in the test.

Tools include, for example, conversion/ capturing tools, and players.

These tools can be in the form of links to commercial tools that can be purchased by the labs and proponents, or source (free) code (compiled Matlab or C).

If possible, the exact tools needed for the test will be decided in an audio conference to be arranged in May by Steve and Mylene and set by Phil. If tools are not available, then specifications for needed tools will be developed.

Changes in the definitions section (see in the document) that were agreed in the reflector and accept editions in chapter 6.

Discussion on the display of qcif sequences:

It was **decided** that no up-sampling or down-sampling is to be used in the player, since this introduces artifacts. The sequences should be displayed in native (maximum) resolution.

Concerning the proposal of having a smaller display for qcif, it was **decided** that the Tools group will recommend a proper size/resolution for the display to be used in the tests.

Editor's note in 6.1.2 – decision to be made by Tools group.

It was **decided** that the Tools mentioned in Section 6.1.2 will be made available by Arthur in the password protected ftp server. BT will take the responsibility to provide the guidelines to use these tools.

The list of companies which will provide sample test sequences was increased (sequences only for research proposes). These sequences will be posted by Chulhee Lee in a new secure ftp server (password protected) on the VQEG site. Advertisement, music, sports, and movies content seem to be needed.

#### Discussion on quality of source home videos:

It was previously decided that any quality of source home video (captured by different types of cameras) would be accepted. Previously, there was an interest in low-quality home videos. But some argued that this might “confuse” the metrics since the range of quality is big and the reference might not be of good quality. The hidden reference clearly eliminates this problem for the subjective test. But, we need to understand what the consequences of this decision are.

It was **decided** that the quality of the source will be considered acceptable if it is judged by an expert as having a visual quality of good or excellent (4 or 5 in an ACR scale). It was left to be decided if the home video should be separated into another category (low quality video?).

Editor’s note on section 6.2.1 – to be filled in later. Changes in 6.3 agreed on Web-conference.

#### Discussion on maximum anomalous frame repetition (Pausing with skipping, Section 6.3.4):

Section 6.3.2 says packet delays are between 100 ms and 5 seconds. It is not specified what the maximum delay is.

**Decision about limits? To be decided later. (6s out of 10s ???)**

#### Outline of open issues (Not on test plan):

New definitions: √

Shall experiments be conducted before or after the model submission? √

Common subset of SRC and HRCs √

Proportion of public, private, and secret experiments

Fees and relationship to proponents conducting and not conducting experiments

Means to exchange experiments and database

Calibration verification of video sequences

## Cross lab experiment design plan

### New definitions:

test material = source, PVS, and experiment data.

public = pool of test material accessible to all proponents prior to submission of models.

private = pool of test material known to one or more, but not all proponents prior to submission of models.

secret = pool of test material that has not been seen by the proponents prior to the submission of models.

### Discussion on "Common subset of SRC and HRCs" and "Proportion of public, private, and secret experiments":

Proponents asked how many PVS they can run.

If the experiment is run before the submission?

If the experiment is run after the submission?

### Proposals:

David's proposal: (current proposal)

Proponents who want generate a set of PVS and send to ILG.

Then ILG selects subsets of these sets of PVS and sends them to the proponents

Proponents and ILG run the experiment using these subsets. (similar to FR test?)

Chris' proposal:

Proponents who want generate a set of PVS and send to ILG.

ILG make these sets of PVS public.

ILG selects subsets of these sets and sends the selection to the proponents.

ILG and proponents run the experiment

(Problem with this approach is that the proponents can train their models if the total number of PVS is not big enough.)

Steve's proposal:

Pay the ILG to run all experiments, PVSs, HRCs, sources, etc.

This approach has the advantage of having "cheat" proof.

Fillipo's proposal: (The goal of having many labs is to have many HRCs and, consequently, a bigger database. )

Have the proponent provide the description of the HRC they are intending to use to VOEG (to be decided if ILG or all). This way we can have an idea of the distribution of HRCs.

Make a subset of sources public (large enough).

Model submission.

ILG selects sources.

The proponents are told which combinations of sources x HRCs to use in the experiments. (similar to Margaret's proposal?)

Margaret's proposal (modified and approved during the meeting)

Source video sequences (e.g., 12-second AVI files containing VGA, CIF or QCIF) are collected and become a public SRC pool. ILG will collect separate secret SRC pool.

Each organization gives a list to VQEG of HRCs they can create.

The initial HRC list to be used for each experiment (by the end of May 2005) is going to be written by Intel.

VQEG decides on a public list of HRC and ILG on a secret.

ILG will agree upon video sequences to be included in every experiment, as proposed by NTT. Up to **10%** of PVSs will be common to all experiments. The set of PVSs will be selected to span a full range of qualities.

Proponents submit their models.

ILG informs organizations what PVSs to be created and send them the SRCs.

ILG creates a set of secret SRCs and secret HRCs. Secret SRC x HRC must be included in the final experiment. In the subjective test approximately **50%** material will be public and **50%** will be secret. (**Exact proportion to be defined**)

The ILG will finalize the designs for each experiment.

Each organization (and ILG) runs the test & submits results to the ILG.

### **Decisions:**

Vote: Training and validation in the same dataset (proportions to be defined) ? NO. (majority)

Decision: Margaret's proposal – rewrite section 6.3. (see above)

The ILG will make available a public source database and generate also a secret source database.

Each proponent is responsible to sending to the ILG a list of HRCs they can generate

An initial list of HRCs will be done by Intel.

Proportions of PVSs common to all experiments is up to 10%.

List of HRCs to be given prior the test.

ILG is responsible for doing the final decision on what PVSs goes in each test.

Back to chapter 7

Filenames in chapter 7 need to be corrected. (will be done later)

Discussion on the output and input formats. (minor changes)

Discussion on need of splitting the RR models in 2 programs (like in the phase 2) to guarantee that the model is really doing what it is supposed to do. It was decided that we will use the RRNR-TV test plan text for this section.

Maximum percentage of data to be checked by the ILG will be 5 PVSs (randomly selected).

Operational systems (workstations for running the objective models) to be finalized.

Discussion on maximum temporal misalignment. From the Study Group 9-12 JRG-MMQA meeting, a maximum of 10% or 1s (which one is lower) should be used. It was decided that this issue has to be investigated before a decision can be made. Information on the subject will be gathered so that this decision can be taken. It is expected that a decision will be made by May 30, 2005. Chulhee Lee will be the point of contact for this issue.

Friday April 29<sup>th</sup>, 2005

Announcement of The First International Conference on Image Media Quality and its application by Mr. Horita from Toyama University (<http://www.ee.kagu.tus.ac.jp/IMQA2005>)

Review of Thursday Minutes

Discussion on "variable delay problems"

Chulhee Lee has distributed an illustration of the problem.

It was **decided** that this issue should be discussed in a tele-conference in May to be organized by Chulhee Lee.

Discussion on Section 7.4 (registration/ calibration)

It was **decided** that all values in the technical criteria (bullets in the beginning of the section) list are to be decided or reconsidered.

It was **decided** that NTIA, NTT, SWISSQUAL, OPTICOM will provide information to base the above decision by the 30th of June, 2005.

It was **decided** that each organization responsible for generating the PVSs is also responsible to check if their PVSs are within the calibration and registration specifications of the test plan. That same organization is also responsible to double check if the PVSs of another organization are also within these requirements.

It was **decided** that if after the experiment PVSs are found to be out of the calibration and registration limits this PVS

should be removed from the data set. ILG will decide if the PVS is outside the limit.

Deleted 7.5.1 and 7.5.2

## Chapter 5

Phil: Report of Display settings and Viewing Environment for MM test

It was **decided** that the viewing environment and display specifications will be decided on an audio conference to be organized by Kjell Brunnstrom.

Section 5.1. – added IRCCyN to the ILG.

Section 5.2. – added Toyama University to the list of test laboratories.

It was **decided** that the initial HRC list provided by Intel should have enough details so proponents know what to expect.

It was **decided** that the specific list of HRCs, (e.g. actual configuration of HRCs, which will be used) will not be distributed to proponents until after models have been submitted.

It was **decided** that the maximum number of subjective experiments run by any one proponent laboratory is 3 times the lowest non-zero number run by any other proponent laboratory, per image size.

It was **decided** that the maximum number of PVSs submitted by any proponent to the overall test should not exceed 20%.

Section 5.3. – Discussion of the Test Schedule

Close of meeting.