

## **Draft VQEG meeting minutes**

### **Kyoto March 3 - 7, 2008**

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 Meeting Minutes**

**Kyoto, March 3 -7, 2008**

### **Monday 3 March 2008**

Monday (Thanks to Phil Corriveau (INTEL) for taking notes.)

Introduction - VQEG

8:30 - 9:00pm availability..

Introductions of each participants

Updates from the Different Groups..

ILG - update - Kjell

Most activities around the MM test.

Test program for MM was completed for the subjective testing

Generated a common set of PVS for the test.

Subjective testing is complete by all labs.

One ILG labs completed the data preparation tool for analysis

Verification and Analysis started.

Report work will involve the ILGs and the rest of VQEG.

Shot some HDTV source from NTIA - sources available for the use of all 1080p

Please share your drives with others to enable copying of files.

That is all for the ILG...

RRNR-TV - Lee

Three proponents

NTIA - NEC - Yonsie University.

Try to finish this off since it has been hanging around for a while.

Materials will be distributed in this meeting.

Only real issue is subjective testing.

Multi-media - Kjell

There have been three audio calls between the meetings.

All the PVS are completed

Ran the subjective testing

All the data is in both Objective and Subjective data - there are issues we will dive into.

Data analysis needs to be completed and will be a large part of this meeting.

The further we get in the report writing in this meeting the better the chance of hitting the standards window - Present the draft rec at the ITU-T SG9 May 5th meeting.

Final report to SG9, SG12, and ITU\_R- 6

Showed the agenda for review - and approval.

HDTV update - Margaret

There is an updated Test plan on the reflector

Review it on the HDTV portion and finalize the plan and set dates for the model submission and test dates for testing.

Tools and Subjective Labs Setup Group (Le Callet)

There is a potential submission - from Yves regarding how MPEG handles for IP

POC Source and HRC collection - Lee

No Update.

Passing the agreements around to be signed so that we are able to get video content shared and distributed

AR - Arthur to provide names of Companies providing what sequences.

Hybrid - Perceptual / Bitstream (Junic/Lee)

Berger - taking over the co-chair of that group..

DECISION - Berger has taken over the role from Pero.

Calibration Verification - SW project (Webster/Lee)

Full reference and Reduced-reference models standardized calibration processing.

Issue is the J.244 had some modification that is editorial or more than editorial and this is to be decided.

More review of the submissions (P.910rev and J.244) to form consensus at JRG-MMQA and

hopefully to get NHK to change their vote from no - yes and overcome their objections

Discussion Wednesday at 1pm....

NTIA – J.cal trial use in the USA and Canada ANSI standard - 18months re-address and approve or change in the standard method. In the APP in SG9.

Project for Collaborative Development (Alex)

Plan was to meet and document the test procedure to be approved at the next meeting.

Intermediate meetings and draft a proposal for the next meeting.

Final Report MM discussion (Volunteers and Assignments)

To Be Determined.

Presentation from Yoshio - Collaboration with IEC TC100 Coordination.

Pioneer Corporation...

DECISION: He will be VQEG Liaison to IEC-TC100

Quality has started in Pioneer

Mainly for Consumer products.

Concerns - AV sync, Loudness in multi-channel, Colour management and measurements and other issues...

TC - Intel should get involved in this activity. TC100

GET SLIDE DECK. (on ftpsite)

Display is the core in the home system and video and audio is very important. Content to end-user.

Receiver - set-top box and compression system for internal storage.

International standards organization are there rules for engagement.

Each country - have formal bodies for standards and they are under the ISO, ITU etc..

Hope that there can be liaison with other groups to move forward the standards work on quality and need technical advice and relationships otherwise, there is a gap.

DECISION - Yoshio will be the liaison between TC100 and VQEG

MM / JRG-MMQA portion - Kjell

Source - Video -

ILG - source receiving scores under the score of 4.0

Uncertain ground - discard should be completed before receiving the data form proponents.

Authority is in question:

1. ILG should be responsible for SRC discard
2. Believe the decision Criteria - Call on the field stands unless there is overwhelming data to change. (set the bar very high for discarding source)
  - a. Visually inspect all - scores are reasonable
  - b. SRC selected and assigned to experiments everything was well balanced
  - c. No visual examination will be justification for rejection of any source material (experts agreed before hand)
  - d. Main idea was to set a bar of 4.0 to avoid compression of the subjective scores.. Limiting the score could be annoying if for many sequences. Not this case
  - e. Analysis on the Common set - elicit trust for the source scores

ILG - Decided not to discard any source and associated PVS's Approved by the VQEG body.

Lowest score for source was - 3.29 - VGA 11 SRC 4

3.38 second.

3.5 - had (2 sources..)

Most are fairly close to 4.0 (3.92, 3.88)

Notes should be made in the final report to understand - (Margaret notes will be distributed)

Understanding if the objective data first then subjective data - and the reverse was true.. Why did we put deadlines.. So that we could avoid debate. Subjective data first then the objective...

Test Lab Report Template - need to have a common format.

Review the Template now....

Decision Remove the Chair and Table from the report.

What level of subjective data should be included in the report?

Decision - Delete from Template the subjective data.

Afternoon.

Validation of Subjective Data.

How many check the entire data set? Is there anyone.

Quan - recomputed the MOS and DMOS values for all experiments and checked them to see if there were any issues and then mailed them out.

And then they were double checked.

Vittorio data is the only ones that have not been doubled check. Released a new version still needs to be double checked.

One of the data sets is missing a data point - suggestion was to put the average into the missing slot for the lab values for the missing data point.

Decision - Average across the Labs will replace the missing data point. Document the details in the final report. C11 and C14 were mixed and thus replace with other lab averages.

Someone from the ILG will go in to the .tab files and fill in the missing value with the averages.. Data needs to be added and deleted. PVS wrongly used -needs to be deleted and then add the average computed from the other subjective tests. Quan says the ILG should do this.

We could get more accurate est. Average and then doing a linear fit for the values that do exist and then allow us to get the est better using the fit calculation. (NTIA) will do the fit. Straight linear optimization using the common set ignoring the missing data point.

NTT Presentation - (Is on the website.. )

If we use condition based analysis it provides more stable data. And allows us to predict things more accurately.

Per condition - you mean per HRC - Eliminate common data points

Eliminating the common set from the analysis.

Take the average of the common set.

Weight of it.

Quan talking about the impartiality of what VQEG is supposed to do - we independently test the models and then if we change the analysis after data collection we lose the credibility of the group and the impartiality of the results.

8.3.3 - proposes another analysis can be supplied.. So the proposal from Juin NTT is valid and this proposal stands.

We at NTIA have a long standing history of per condition analysis - and things are different from the stimuli etc..

Freeze frames for the HRC's - The PVS was out before the deadline - they were checking for calibration. Different routines for detecting source and thus we could not id such an error.

The problem is the Source - violated the test conditions and the codec was not a valid one.

Corresponding section - there is no restriction to the bit rate - very usual condition.

Margaret

Ambiguous if this HRC is allowed in the test plan - sees both sides of the argument.

Accident was not what was intended.

Bigger issue whether we can toss an HRC at this point of time. We need to justify why we discarded and that it is not done to make a model look better. Many organizations feel that discarding HRC after both the subjective and objective data have been received. (she agreed it was an unauthorized change and we did not predict for it)

Chulee lee - unauthorized change of the test design.

Marcus - PVS is not coded in any way - lossless - not lossless since there are frame freezes.

Exclude one HRC - due to non-conformance opinion only.

KDD - Agree to remove

NTT - Better to remove.

Marie - DT - is it really not realistic. - only case where there is no compression and freeze.

Kwill - Remove the sequence since it violates the test plan - Filippo - Gray area.

NTIA - Keep the HRC since we have all the related data and they should be included in the results.

Ericson - Keep the HRC - could be there..

Swiss-Qual - Keep the HRC

Alex - personal problem - could this be a decision from the ILG - and it eliminates the impact on the data.

BT - Disagree with removing the data.

Quan - Does not see the point in continuing arguing over the point - would join Marcus - Once the data is available it is not a good thing to change our minds. (Alex makes a good point).

We should wait for all the data before the final report - waiting for France Telecom.

Now working on the report for the MM test results.

Draft is the production of David Hands - took the FRTV II report and edited it for the MM final report.

Split into subgroups to work on the MM report - ( sections )

Reconvened at 5:00PM. Margaret Pinson (NTIA) discussed some points regarding the proposed removal of the Common set of sequences in the data analysis. She will send an email to the mctest reflector.

## VQEG Minutes from Tuesday March 4, 2008, in Kyoto

Thanks to Margaret Pinson (NTIA/ITS) for taking the notes.

France Telecom (FT) subjective data is available. FT data must be re-formatted.

NTT withdraws proposal to remove common set from primary analysis (i.e., per-clip analysis).

HRC analysis is allowed in test plan as secondary analysis. Issue is whether to include common set in the HRC analysis.

**Agreement was reached** to include HRC analysis as secondary analysis in the MM final report. [Note: test plan indicates this is optional].

**Agreement was reached** to remove common set from this HRC analysis (i.e., because there is only one SRC for each HRC, HRC analysis is not possible on the common set).

**Agreement was reached** that this difference between the treatment of the common set in per-clip analysis and HRC analysis must be clearly stated in the final report.

**Agreement was reached** that the number of data points for each fit (16) must be clearly stated in the final report.

- Proposal was made to map all tests to one scale using common set, then do HRC analysis on all HRCs simultaneously. This will be discussed later.

Discussions continued on V08 HRCs 7, 8, & 9. Introduction & minutes from Seoul 2004 meeting both indicate hard upper bit-rate limit of 4 mbits/s, and these HRCs are clearly above that limit.

For removal: KDDI, NTT, NEC, KWILL, Symetricom, NTIA, Ericson, Yonsei, Opticom,  
Against removal: Swissqual.

**Agreement was reached** to remove V08 HRCs 7, 8, & 9 from primary analysis.

For including secondary analysis of V08 with these three HRCs for all models: Acreo, IRCCyN, Intel, NTIA, Psytechnics, Swissqual (6).

For excluding: KDDI, NTT, NEC, Symetricom, Ericson, Opticom, Yonsei, CRC (8).

**Agreement was reached** to exclude analysis of V08 with HRCs 7, 8, & 9 in another section of the report.

**Agreement was reached** to include in the final report an explanation of what occurred with V08 HRCs 7, 8, 9 and other changes to raw data.

**Agreement was reached** that two Excel files will be made available to all proponents and ILG. The first Excel file would have all MOS scores (i.e., as produced by subjective tests), before any PVSs are removed. The second Excel file will have the final data used for analysis (e.g., after V08 HRCs 7, 8, & 9 are removed). Only the second Excel file will be in the final report.

- Discussions will be needed on the use of VQEG data in papers and other reports, in particular the subjective data from MM experiment. This will be discussed later.

**Agreement was reached** to *not* map all tests to one scale using the common set, then do HRC analysis on all HRCs simultaneously. HRC analysis will be performed on each test separately.

Discussion ensued on procedure for selecting fitting coefficients.

**Agreement was reached** for the missing C11 & C14 common set values (one PVS each) to use the average of other data sets' values for these PVSs. The ILG will be responsible for this. That is, take the average of MOS, DMOS, and Stdev.

Motivation: the improvement from a linear fit would be minimal, and all CIF data is not yet available.

Other HRCs have been located that violate the MM test plan. These HRCs were listed correctly in the test designs, and no objections were raised prior to the pre-subjective testing deadline. This may be an issue of an old test design, where the HRC used does not match the listed HRC.

**Agreement was reached** concerning V13 HRC 16, which uses a bit-rate that is higher than the bit-rate specified in the MM test plan. This bit-rate was correctly listed in the test design. No objections were raised to this HRC prior to the proponent deadline, so it will be kept. This discrepancy will be noted in the final report.

**Agreement was reached** to use the mapping computed for primary analysis on each model / experiment pairing, will also be used for all secondary analysis (e.g., HRC analysis).

In favor of SRC analysis: NTIA

Opposed: KDDI, NTT, Ericson, Yonsei, BT, Opticom

**Agreement was reached** to not include SRC analysis in the final report.

## Minutes from VQEG the Meeting, Morning Wednesday 5 March 2008

Thanks to Jorgen Gustafsson (Ericsson) and Carolyn Ford (NTIA/ITS) for taking notes.

### Definition of outliers

The definition of the outlier calculation is ambiguous. The definition has previously been change, and can be interpreted in two ways.

**Decision:** The decision was to change the name of the two definitions in the test plan. The name of the top one will be changed to “outlier ratio using standard deviation of the mean” and the other will be changed to “outlier ratio using the standard error of the mean”.

### Information

- There will be a cultural event and a dinner arranged on Thursday.
- An ETSI workshop will be held in Prague about Multimedia Quality assessment.

### Analysis of MM Test Results

Statistical analysis for the MM test was presented by Filippo. There are three ways of calculating the confidence interval in the presented tab file:

- Polynomial No Monotonic Constraint
- Standard error of the mean
- Constrained (calculated with Optimap, in line with the test plan)

Christian provided another Excel file containing statistical data and charts with confidence intervals, RMSE and outlier ratio for the MM models (charts only for the MM models), based on Optimap.

The data files were uploaded to the VQEG ftp server (for the Kyoto meeting). The data is not the official ILG data analysis, but no major changes are expected.

The meeting when unusually quiet when the data was presented...

Jens pointed out that the NR models got significantly lower correlation when using Optimap than the polynomial with monotonic constraint confidence interval (in the data provided by Filippo). Filippo will do some more statistical analysis to double check the correlation values.

There was a discussion about what data to put in the final report. At least the graphs presented by Christian should be included in the report.

Chulhee suggested that the final report should contain enough information to be able to validate an implementation of the standardized model(s). However, this requires that the test sequences are open source.

There was a discussion about what to include in the final report. Arthur said that the MOS and DMOS values from all subjective tests should be included in the final report. Chulhee brought up

the issue if someone would like to use this data for any kind of analysis, will they be allowed to do that? Filippo said that including only the graphs makes the data more difficult to use for others.

Arthur brought up a number of important questions:

- Are the models significantly better than PSNR?
- Should we use the average correlation over all tests?
- Are some models not significantly different?

Christian presented the average correlation values (per video format, taken from the Opticom Excel file). For the FR models Psytechnics and Opticom are the two best, followed by NTT and Yonsei, followed by PSNR, and then the others.

## MM Final Report

A first version of the final report was uploaded to the FTP server and to the Kyoto meeting directory. The file name is VQEG\_MM\_Report\_Final\_v1.doc. Thank you all authors!

KDDI requested that the KDDI results will be removed from the final report.

**Decision:** It was decided to remove the KDDI results from the final results and also from the calculations and data discussed at the meeting. Two sets of data will be saved on the secure MM ftp site: One including the KDDI results and one not including the results.

The goal is to have the MM final report ready before May 3<sup>rd</sup>. The SG 9 meeting starts at May 5<sup>th</sup>.

The final selection of proponents' fitting coefficient will be selected in two steps:

1. The proponents will, before March 14, send in coefficient to Chuhee
2. At March 21 the proponents will make a final selection of what coefficients to use for each test.

At this point, the note taking has been taken over by Carolyn Ford, NTIA/ITS  
It is alternatingly sunny and snowing in Kyoto.

Discussion of the revision P.910 and the use of labels (numbers).

Alexander (DT) presented a paper that studied this issue (on the VQEG website).

Experiment is proposed to study the effect of numbers/no numbers

Arthur will write something up about the issue and put in the JRGMMQ report and present to NHK.

Discussion of J.255 (calibration). Agreed to let NHK have further review.

Carolyn presented a brilliant overview of proposed test method P.trv (Target Recognition Video).

Proponents are reminded that written descriptions of their models need to be available before the SG9 meeting. It will be determined later what the options are for delivery.

The snow has stopped.

Chulhee proposed wording for guidelines for the use of the VQEG MM Phase I data. It will become a statement at the front of the draft final report.

Marcus requested the use of the MM Phase 1 data for use in his PhD dissertation. No objections were raised, it falls under item 3 in the guidelines.

Jens asked who has the IP on the subjective score database.

The snow is back.

Chulhee has contributed a list of contributors to the subjective data, and guidelines for use of the data for the writing of papers. The document will be on the website under the MM project folder, and sent to the MM reflector. The text is included as the last page of these minutes. See attached.

Jens announced that SwissQual is no longer a proponent for FR but is still for NR.

Metrics tools are available on the website.

Review of schedule:

Draft final report will be worked on this week, finished over a teleconference. Audio call on March 27 to decide how to proceed.

Stayed late and divided into small groups to work on sections of the final report.

Marcus was tasked to write sample acknowledgments. Will be included in papers. Here they are:

Begin text from Marcus:

For papers which propose a new model:

Acknowledgement

This publication is [partly] based on the subjective scores collected by the Video Quality Experts Group (VQEG). The results presented in this [paper/publications] are not to be compared to the results presented in the VQEG Final Report of Multimedia Phase I [reference\_goes\_here] because the models in the report were validated using this data. Thus, the data was not available to the models that were submitted to the VQEG evaluation.

For papers which use the subjective data for some kind of analysis:

Acknowledgement

This publication is [partly] based on the subjective scores collected by the Video Quality Experts Group (VQEG). [The/Some of the] subjective experiments used in this study were conducted by several contributors as listed in the VQEG Final Report of Multimedia Phase I [reference\_goes\_here].

End text from Marcus

The executive summary was drafted and added to the draft final report.  
It is not approved at this time. It will be readdressed later Thursday or Friday morning.

Below was discussed, but not agreed.

Begin not agreed text:

### **Issues**

- Use of subjective and objective data of copyrighted materials for publication
- Additional use copyrighted materials (e.g., research, paper publication)
- Contributors of subjective data: ILG design, PVS generation, subjective test, provider of source materials (acknowledgement)

### **Options**

- (1) Includes all contributors as authors if the subjective data is not published. If an author(s) contributes significant new works, they are entitled to make the author order.
- (2) Provide proper acknowledgement to all contributors.

#### **Scenarios.**

- (1) If a paper is written using only the available subjective (objective) data, all the contributors of the subjective data included in the paper should be listed as authors.
- (2) If new data constitutes at least 50% of the data used in the paper and the main conclusion can't be drawn from the existing subjective and objective data, acknowledgement of the contributors would be sufficient.
- (3) If a paper is submitted in 2008, all the contributors of the subjective data included in the paper should be listed as authors.

### **(Note)**

1. If some of the subjective data is published (e.g., Recommendations, VQEG report in the public domain), everyone is entitled to use them with proper citation.
2. If a paper(s) mainly consists of VQEG MM results (e.g., subjective data analysis), all contributors should be included with ILGs on good seats.
3. Each proponent is entitled to publish papers which present their models with proper acknowledgement to all contributors.
4. If a paper uses objective data of other proponents, their permission is required.

**Potential paper topics using the subjective data**

1. Subjective data comparison across cultures (countries, languages)
2. Effects of descriptive adjectives
3. Differences between genders
4. Age effects on subjective data
5. etc

**Potential paper topics based the subjective data, which require additional works**

1. Comparison ACR-HRR with numbers
2. Comparison with DSCQS
3. etc

End of not agreed text.

From here down is agreed to:

Begin agreed text:

An organization that has access to the subjective data and plans to write a paper using the MM subjective data of the other contributors (ILG which designed the tests, organization that generated the PVSs, organization that performed the subjective tests) are asked to inform the contributors about the paper (abstract, goals, etc) and invite them to be co-authors by sending an email to the MM reflector. The co-authors are expected to make a significant contribution in writing the paper (data analysis, experiments, writing, etc). If there is no response to the invitation within 4 weeks, the organization ~~is will proceed allowed~~ to write the paper.

List of the contributors of the subjective data

**Proponents**

~~KDDI~~  
NTT  
Opticom  
Psytechnics  
Swissqual  
Symmetricom  
Yonsei

**ILGs**

Acree  
BT  
~~CRC~~  
Nortel  
~~Ericsson~~  
Intel  
FUB  
NTIA/ITS  
Verizon  
IRCCyN  
FT

**Others**

KDDI

End agreed text

Formatted: Font: Bold

**Minutes VQEG meeting  
Thursday March 6, 2008**

Thanks to Yves Dhondt, of Ghent University for taking the notes.

**Start**

Approval of Wednesday's revised minutes.

**HDTV (09:30 – 11:30)**

Resolving issues in the HDTV test plan version 2.0:

- 1) Do we want to restrict the test plan to digital impairments or should analog impairments be included as well?  
Some organizations have interest in analog impairments as well so they will be included.
- 2) Proposed model submission date (including one subjective data set) changed to approximately one year from now.
- 3) The basic software interface for HDTV for input and output data formats will be the same as the one for MM.
- 4) It was agreed to only accept the following formats: 1080i@50, 1080i@60, 720p@50, **720p@60**
- 5) *Original proposal for purchasing HD source by proponents:*  
The company and price should be made public (not the source itself) and at least one member of the ILG should be able to buy the source.  
*Change requested by Kjell:*  
Proponents who buy source should also buy a copy for the ILG. – **accepted.**  
*Remark by Quan:*  
What if a proponent can buy the source for everyone? That is not an issue here; the proposal only covers the case where this is not possible.
- 6) *Changes to the test design*  
Discussion on the number of SRCs per test: 4 organizations want 8, 4 organizations want 10. It was decided to pick 9 SRCs.  
Discussion on the number of PVSs per test: fixed to 162 PVSs (including hidden reference) for every test.  
**Decision:** each test will consist of 162 PVSs (9 SRCs, 18 HRCs (including hidden reference))
- 7) Discussion on the display to use. The display will do some kind of de-interlacing.  
Suggestion 1: everyone should use the exact same monitor: 0 in favor.  
Suggestion 2: the video should be de-interlaced before sending it to the monitor: no voting  
Suggestion 3: everybody tells others which monitor they will use: no voting

Due to the complexity of the subject, and the current lack of time, a subcommittee should be formed which can come up with a proposal regarding monitor specifications. Volunteers for

this group are: Chulhee, Patrick, Marcus, Alexandre and Kjell.

- 8) Subjects should not be monitored during the test. (Removal of paragraph from section 4.8)
  - 9) The proposed text for section 5.3 is accepted without comments. (see v2.1 of HDTV TestPlan)
  - 10) The proposed text for section 5.7 is accepted with the following remarks:
    - for point 1: the frame rate should be fixed
    - for point 2: the frame resolution should be fixed(see v2.1 of HDTV TestPlan)
  
  - 11) The proposed text for section 6.2 is accepted without comments.
  - 12) The proposed text for section 6.3 is accepted with the following comments:
    - a one digit version number should be attached to prevent the problems VQEG had in the MM tests
    - SRC numbers should consist of two digits (01, 02, ...)
    - HRC numbers should consist of two digits (01, 02, ...)For example: xyz\_src01\_hrc04\_v1.avi
  - 13) Discussion about the introduction of section 7.2 on the error conditions. Change of “These error conditions may include, but are not be limited to, the following” into “These error conditions include the following”
  - 14) **Decision:** Inclusion of H.264 (SVC) into the allowed codecs – no objections.
  - 15) Frame freezing up to two seconds is allowed.  
Frame skipping is allowed.  
Limitations are defined in the test plan under section 7.2.4.
  - 16) Vote on the allowance of rewinding as an error: yes: 3 / no: 9.  
**Decision:** Rewinding is not allowed.  
Rather than removing the section concerning rewinding, the test plan is adjusted to make clear that rewinding is not allowed. The appearance of rewinding should be checked visually (not frame by frame).
  - 17) There is an interest in changing the frame rate from 24 to 5 or 10 frames per second (for example for security cameras).  
Would you like to see security cameras supported in this test plan: yes: 0.  
Would you like 24 frames to be the minimum if you can set the frame rate of the encoder manually: yes: 13, no: 1 (lower fr, to 10).
  - 18) Up- and downsampling is allowed as a pre-processing technique.
- RRNR-TV (13:00 – 13:30)**
- 1) NTIA can not cope with the current schedule. It is suggested to give them more time. The date for HRC creation is moved up to May 15.

- 2) The change of date allows K-WLL to join. They will look into it if they can still generate HRCs and run tests. It is suggested that if they can not run any subjective tests, they should hire an European ILG to do it for them. This solves some of the concerns NTIA has.
- 3) Are there any others interested in joining the RR tests? None.
- 4) In a few days, an audio call will be held to take definite decisions.

**Hybrid – Bitstream** (13:30 – 16:30)

- 1) Summary of the decisions made during the Ottawa meeting.
- 2) Presentation by NTT:
  - multiple decoders for each 'codec' should be used
  - how to practically test the previous point: test all, use a regression function based on a base decoder
- 3) Presentation by Ericsson:
  - usage of an offline simulator
  - models operating inside the network (with return channels)
  - rebuffering (mostly for low bit-rate mobile devices, so not for HD)
- 4) Discussion regarding the PVS length based on the inclusion of PVSs:
  - If you want to include rebuffering, PVSs should be longer.
  - If you have freezing without skipping or rebuffering, the PVSs should get longer.
- 5) Do we want freezing without skipping for QCIF, CIF and QVGA? yes: 9 / no: 0  
Do we want freezing without skipping for SD and HD? yes: 1 / no: 10  
**Definition:** Rebuffering is freezing for longer than 0.5s without skipping.
- 6) Based on the previous point:  
For SD/HD how long should the length of PVSs be?
  - a) 10 seconds: 6 (third vote: 5) (second vote: 5) (first vote: 5)
  - b) 16 seconds: 7 (third vote: 5) (second vote: 5) (first vote: 3)
  - c) 30 seconds: (second vote: 0) (first vote: NA)  
For QCIF/CIF/QVGA do you want the length of the PVSs to be fixed?  
Variable: 7  
Fixed: 3  
  
For QCIF/CIF/QVGA with rebuffering (variable length), how long should the length of PVSs be?
  - a) 30s SRC (30s-45s PVS): 1 (first vote: 3)
  - b) 16s SRC (16s-24s PVS): 7 (first vote: 3)  
For QCIF/CIF/QVGA without rebuffering (fixed length), how long should the length of the PVSs be?  
10 seconds
- 7) Subjective testing should be using ACR with HRR  
Graphical, continuous scale might be used

No decision on numbers at this moment.

- 8) Model types:  
Proposal: P.NAMS should get information from the client (rebuffering)  
  
Hybrid models  
Model A (P.NAMS): TS, RTP, UDP, IP headers & (optional longer freezing)  
Model B (P.NBAMS): TS, RTP, UDP, IP headers, ES information & (optional longer freezing)  
Remark: ES information = bitstream data
- 9) In his slides Chulhee made a list of proponents and the preferred input types for the different models.
- 10) Reference decoders:  
for H.264: JM reference software  
for MPEG-2: FFMPEG  
for MPEG-4: not decided yet, Momusys needs checking first
- 11) Capture file format:  
PCAP  
RTPdump  
No decision is taken.
- 12) Possible streaming servers:  
VLC  
Helix  
Darwin  
Live555

**HDTV** (16:45 – 17:45)

- 1) Presentation by IRCCyn  
Topic: Removing 720p from the test.
- 2) Based on the previous point, the following proposal was made:  
Eliminate 720p as a separate testing, and include 720p into 1080i tests? yes: 11 votes / no: 2 votes

How might be interested in being a proponent for the HD plan? K-WLL, NEC, NTT, KDDI, SwissQual, Yonsei, Opticom, NTIA, Symmetricom

Preferred source:

- 1080i only: 3 votes
- 1080i / 1080 p: 6 votes
- 1080p only: 2 votes

Who would like to limit each experiment to just have 1080i or 1080p (you cannot mix them in one test)? 10 votes in favor.

Is there anyone who wants to combine 1080i and 1080p in one test? 0 votes in favor.

- 3) Presentation by IRCCyn  
Topic: Comparison of ACR and SAMVIQ.

**Friday Minutes**  
**March 7, 2008 Kyoto VQEG Meeting**  
**MM Discussions Continued from Wednesday**

Thanks to Margaret Pinson (NTIA/ITS) for taking the notes.

On the issue of Outlier Ratio, minutes from previous meetings were investigated. At the Scottsdale, Arizona, USA meeting April 25-29, 2005, the following agreement was found:

**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).

That is, a contribution at the Scottsdale meeting proposed a change from standard deviation to standard error. This proposed change was rejected. The presence of standard deviation in version 1.19 of the MM test plan is clearly an editorial error.

**Agreement was reached:** MM test plan will use standard error in equation 11 for Outlier ratio (i.e., divide by the square root of the number of viewers).

**Agreement was reached:** Additionally, 1.96 will be used instead of 2.0. (Note: this is mathematical fact, the 95% CI cannot be computed with 2.0).

**Agreement was reached:** to update the test plan (i.e., produce a new version) to avoid this error occurring in the future.

**Announcement:** Symetricom is withdrawing their Full Reference CIF model from the MM test

Model Descriptions Written & Submitted for MM Final Report

- Opticom yes
- NTT yes
- Swissqual yes
- Psytechnics yes
- Yonsei yes

Note: withdrawn models will not be described

Lab Templates Written & Submitted for MM Final Report:

- Swissqual yes
- CRC yes
- FUB missing
- Acreo yes
- FT missing
- IRCCyN missing, March 12
- Opticom yes
- KDDI missing, March 10
- NTT missing, March 12
- Psytechnics missing, perhaps March 19
- Yonsei missing
- Symetricom March 19
- Verizon missing

Greg and Filippo and Patrick will be in charge of data analysis.

The draft executive summary was examined. See [ftp://vqeg.its.bldrdoc.gov/Documents/VOEG\\_Kyoto\\_Mar08/MeetingFiles/MMFinalReportStuff/VOEG\\_MM\\_Report\\_Final\\_v1.1.doc](ftp://vqeg.its.bldrdoc.gov/Documents/VOEG_Kyoto_Mar08/MeetingFiles/MMFinalReportStuff/VOEG_MM_Report_Final_v1.1.doc) Notes were added showing where some text needs to be re-addressed when final results are available. See updated version for changes.

Chulhee's email formally objecting to aggregation by way of ranking models' performance was discussed.

Drafting Group for MM Draft Recommendations:

Arthur, Akira, Phil, David, Alex, Christian, Kjell, Filippo, Patrick, Jörgen, Chulhee, Jens, and Quan

**Agreement was reached** that some FR models examined by the MM test performed well enough to be included in the normative sections of Recommendations. The scope of these recommendations should be written carefully to ensure that the "use of the models" is defined appropriately.

**Agreement was reached** that the MM final report that some models were withdrawn. No proponent names will be mentioned.

**PSNR:** NTT PSNR uses a sophisticated temporal registration algorithm, which makes PSNR look "better" than it is typically. NTIA's PSNR was discussed as an additional PSNR to include in the report.

**Note:** VQEG would like to see NTIA's PSNR results and include them in the MM Final Report.

**Agreement was reached** that some of the RR models may be considered for standardization. The scope of these recommendations should be written carefully to ensure that the "use of the models" is defined appropriately.

**Agreement was reached** that the proposed text "some QCIF NR models may be appropriate for standardization" should be excluded from the executive summary for now. (6 voted include, 7 voted exclude) This issue will wait for supplementary analysis.

**Agreement was reached** that no CIF or VGA NR models may be appropriate for standardization. Secondary analysis will be required to decide on this issue. The scope of these recommendations should be written carefully to ensure that the "use of the models" is defined appropriately. No mention will be made of informative annexes. Partial text will indicate an upcoming supplementary analysis summary.

HDTV test plan edited further. Several agreements were reached (e.g., 25fps and 30fps specified for 1080p, rules by which PVSs may be discarded after model submission, calibration limits). See updated version of HDTV test plan.

**Agreement was reached** not to include 24fps in HDTV test plan.

**Agreement reached** to add to HDTV Test Plan:

- date by which proponents state intention to submit model.
- At this date, format must be specified (e.g., 1080i 25fps)
- At this date, proponents must specify whether they intend to include 720p HRCs
- We desire that all 4 formats be included – 1080p & 1080i, 25fps & 30fps (i.e., at least one test at each format, and 720p HRCs included)