

Committee T1 Performance  
Standards Contribution

106

Document Number:  
TIBBS File:

T1A1.5/98-206 + G-2.1.6/82

DATE:

July 13, 1998

STANDARDS PROJECT:

Analog Interface Performance Specifications for Digital Video  
Teleconferencing/Video Telephony Service (T1Q1-12)

SUBJECT:

ITU VQEG Activity Report

SOURCE:

NTIA/ITS

CONTACT:

Arthur Webster (303-497-3567)

KEY WORDS:

video quality, picture quality, subjective, objective,  
performance assessment, multimedia, JRG.VQEG

DISTRIBUTION:

Working Group T1A1.5

# ITU VQEG Activity Report

The Video Quality Experts Group (VQEG) met in Gaithersburg, Maryland at the NIST facilities May 27-29, 1998. The report of that meeting is attached as ANNEX 1 (page 2 of this contribution). Final decisions were reached at this meeting to allow the validation testing for objective methods of video quality assessment to proceed as planned.

## Summary of Activity

- It was agreed that Arthur Webster of NTIA/ITS and Philip Corriveau of CRC will be the Co-Chairs of VQEG.
- Test Scenes were selected (by the ILSC): 10 each at 50/525 and 60/625.
- The final list of HRCs was agreed upon.
- Numerous details concerning the Subjective and Objective test plans were finalized.
- A Schedule was established for the work and can be found on page 6 of this contribution. The final results of the test are expected in February, 1999.

## Final List of Proponents in the Validation Test:

Tektronix / Sarnoff USA

CPqD-TELEBRAS Brazil

Institut fuer Nachrichtentechnik - Germany

KPN Research The Netherlands / Swisscom CIT

NTIA/ITS USA

KDD Japan

NASA USA

TAPESTRIES Europe

NHK (Japan Broadcasting Corporation)/ Mitsubishi Electric Corp. Japan

Swiss Federal Institute of Technology (EPFL) - Switzerland

## Future Work of VQEG:

- A number of areas were identified as possible candidates for future work of VQEG. It was agreed that the next effort will involve objective quality measurement of:
  - Low Bit Rate Video (16 kb/s up)
  - Transmission Errors
  - Compressed Source Measurements (aka In-Service)
- Other areas of future video quality work identified were:
  - 5 second Scene Length test
  - Testing using multiple viewing distances (e.g. 3H and 6H)
  - Still Image transmission objective measures
  - HDTV objective measurement
  - Sub-Threshold objective measurement
  - Combining sub-threshold measurements to predict quality
  - Audio-video synchronization & Audiovisual quality
  - Threshold visibility
  - Continuous quality evaluation

## **ANNEX 1: Report of the Gaithersburg VQEG meeting**

### **Decisions made at the VQEG meeting at the NIST facility Gaithersburg, Maryland May 27-29, 1998**

#### **CO-CHAIRS OF VQEG**

It was agreed that Arthur Webster and Philip Corriveau will be the Co-Chairs of VQEG.

#### **ILSC REPORT**

Test Scenes were selected: 10 each at 50/525 and 60/625.

→ Volunteers are needed for HRC processing--especially in the 525 half of the test. If help is not found to run the 525 HRCs, the test will proceed with only 625 Scene/HRC test conditions.

→ Funds and/or tape stock and/or labor is needed from the proponents to help get the subjective viewing tapes edited.

A finance committee was established to coordinate the financial aspects of the test:

Vittorio Baroncini  
Phil Corriveau  
Stephane Pefferkorn  
Laura Contin  
Arthur Webster

#### **DATA PRIVACY ISSUES**

On all public documents relating to this validation test:

Subjective laboratories will be identified by numbers.

Objective models/proponents will be identified by letters.

The only results of the objective models to be shared publicly are the prediction results.

Coder-decoders will not be identified by manufacturer.

#### **METRICS**

ANOVA, Kurtosis, and weighted RMS error were eliminated as metrics for objective model evaluation.

The following metrics will be used for objective model evaluation:

- Simple RMS error between model output and the raw (each viewer's difference) subjective scores.

- Pearson correlation coefficient on both the DMOS and the raw subjective scores.
- Spearman rank-order correlation.
- The outlier measure in the current version of the objective test plan.

Regarding the use of complexity as a criterion for algorithm acceptance:  
The text will remain as it stands in the Objective Test Plan.

#### LAB TO LAB COMPARISONS

VQEG will adopt lab-to-lab comparisons as a benchmark in the evaluation of objective measurement methods.

#### INTELLECTUAL PROPERTY ISSUES

→ It was agreed that each proponent will send to the reflector a statement that they intend to abide by the ITU patent policy if their algorithm is selected for inclusion in an ITU Recommendation.

This patent policy for the ITU-T can be found in document Resolution 1 at the ITU Web site. ([http://ties.itu.int/itudocr/itu-t/wtsc-96/res/001\\_e\\_51047.doc](http://ties.itu.int/itudocr/itu-t/wtsc-96/res/001_e_51047.doc))

Accessing this requires a TIES account. If requested I can excerpt the pertinent passages and place them on the ituvidq ftp site.

For the ITU-R I expect a similar document can be found. I would ask someone from ITU-R 10-11Q if they know where this is to be found.

#### ALGORITHM DISCLOSURE

It was also agreed that for those proponents that are selected for recommendation to the ITU Study Groups for Recommendation, a description of the algorithm is required that will allow full implementation. This could include source code but source code is not required. ITU Study Groups make their own decisions, but it is VQEG's position to suggest that source code will not be necessary.

#### NON-DISCLOSURE AGREEMENT

A non-disclosure agreement was drafted and accepted. Members of the ILSC will read and agree to the terms. No signatures will be taken—the honor system will be used. The document is to ensure that ILSC members are aware of the information needs to remain confidential. Any persons who, by helping the ILSC to conduct their work, should learn of the test sequences or the HRCs shall become members of the ILSC and must read the non-disclosure agreement.

#### INSTRUCTIONS TO VIEWERS

Left to the ILSC.

#### → RANDOMIZATIONS

Two, if resources permit--one otherwise.

## TEST SCENE LENGTH

8 seconds shown to the subjects. 10 frames will be allowed before and after the 8 second viewing material to allow the codecs to settle down. One second before and after the test sequence will contain the alignment markers provided by Tektronix. The marker data will be saved to allow for data analysis and verification. The marker data may be removed before the objective models process the data or some other method will be devised to ensure that the models do not utilize any information contained in the marker.

## TIME SCHEDULE (see attached table)

All proponents must send to the reflector by June 22, 1998 a statement of their intention to submit a model. With this, a statement must be included stating that they will abide by the ITU patent policy if their algorithm is chosen for inclusion in a Recommendation.

By July 22, 1998 an executable code must be delivered to these two ILSC sites:

CRC  
CSELT

The code (or preliminary test code) can be delivered before this date to your independent laboratory to work out any software compatibility problems. For the location of your assigned objective lab please contact:

Phil Corriveau  
at CRC, Canada  
tel: (613)998-7822  
fax: (613) 998-7823  
eMail: philc@dgbt.doc.ca

All code must be working by August 7, 1998. No further modifications will be allowed beyond this date. It is recommended that proponents begin working to get their programs running at their assigned objective lab as soon as possible.

## OTHER DECISIONS ON SUBJECTIVE TEST PLAN

### Validation of Monitors:

Monitors will be checked for resolution, alignment, and chromaticity. Subjective labs will report make and model of monitors to be used for tests before July 1, 1998.

### Test conditions:

- The ILSC will seek to have at least 5% but no more than 10% of the conditions be perceptually transparent at the viewing distance.
- All test sequences will be shown full size (no letterbox).
- All reference (source) will be ITU-R 601 format.
- No noisy source scenes will be used.
- Normalization of processed sequences will be as specified in the Objective test plan and further detailed in Document VQEG013.
- Viewing Distance: 5H

## HRC LIST (see new list attached)

768 in.

VHS out.

Transmission Errors in.

The subjective test will be split into high and low quality level tests.

The ILSC will try, as a guideline, to choose codecs from several manufacturers.

If resources are not found to run HRCs and edit the viewing tapes, some HRCs will be eliminated from the test. The ILSC will decide which HRCs will be deleted if it becomes necessary.

## DECISION MAKING PROCEDURES:

The chairs note that the VQEG is an unofficial group of experts meeting to share expertise and resources for the primary goal of expediting technical work leading to the creation of ITU Recommendations. As such it has no rules other than those to which it agrees. The following rules were agreed to at the meeting.

1. In coming to decisions, the chairs will always seek to achieve consensus.

Voting is undertaken only as a last resort when no more time is available to work out a consensus.

2. To change a previous decision 2/3 of present voting participants must agree.

3. Turin decisions are considered "previous decisions".

4. One vote is allowed per organization present at a meeting.

5. To make a new decision simple majority of those voting is required.

6. Chairpersons are allowed to vote.

Jean-Pierre Evain is opposed to chairpersons voting.

## FUTURE WORK

A number of areas were identified as possible candidates for future work of VQEG. It was agreed that the next effort will involve objective quality measurement of:

Low Bit Rate Video (16 kb/s up)

Transmission Errors

Compressed Source Measurements (aka In-Service)

(a better term for this must be found as "in-service" measurements can be made in some situations with the full Source video available.)

The other areas of future video quality work were:

5 second Scene Length test

Testing using multiple viewing distances (e.g. 3H and 6H)

Still Image transmission objective measures

HDTV objective measurement

Sub-Threshold objective measurement

Audio-video synchronization

Audiovisual quality

Threshold visibility

Continuous quality evaluation

Combining sub-threshold measurements to predict quality

## Schedule

Action	Who	Dead-line
Sending patterns for the normalisation to CCETT & CRC	Tektronix	12 Jun 98
Proponents declare intention and submit patent policy agreement.	All Proponents	22 Jun
Adding patterns to the source sequences and sending them on D1 tapes to HRC processing sites	CCETT & CRC	3 Jul <sup>1</sup>
Executable code to objective labs and ILSC Chairs	All Proponents	22 Jul
Final, working executable code to objective labs and ILSC Chairs	All Proponents	7 Aug
HRC processed sequences and 'patterned' source sequences on D1 tapes to Tektronix for normalization	IRT, RAI, others (TBD)	8 Aug
Normalized D1 material to the editing sites	Tektronix	11 Sep
Normalized source and encoded material on Exabyte (2 Gbytes) tapes to the proponents and objective sites	Tektronix	28 Sep
Normalized source and encoded material on DAT tapes to some of the proponents and objective sites	NTIA	9 Oct
Editing of the test tapes done and sent to the subjective test sites	FUB (?), CCETT(?), CRC(?), NTIA (?)	9 Oct
Subjective tests complete.	ATTC, Berkom, CCETT, CSELT, CRC, DoCatA, FUB, RAI, Teracom	13 Nov
Objective test complete.	ATT & NIST(SGI) FUB&CRC (Sun) IRT (PC)	11 Dec
Individual Labs Statistical analysis of subjective test data complete	CRC, (CSELT), CCETT, NIST	11 Dec
Discussion of results of subjective tests & release of subjective data to the proponents and whole of VQEG	ILSC	4 Jan 99 <sup>2</sup>
Analysis of 'correlation' between objective and subjective data completed.	NIST	5 Feb
Meeting at FUB in Rome to discuss results and the preparation of the final report	VQEG	TBD Feb or March

<sup>1</sup> In case some of the HRC processing sites are proponents, the source sequences will be delivered by the 7<sup>th</sup> of August, instead of the 3<sup>rd</sup> of July and a one month delay will be introduced into the rest of the schedule.

<sup>2</sup> Considering Christmas holidays, it's maybe better to move this dead-line to the 11<sup>th</sup> of January and the correlation analysis to the 12<sup>th</sup> of February [LC]

**HRC LIST (new numbers)**  
**Shaded area is overlap between Low(A) and High(B) tests.**

	A	B	BIT RATE		RES	METHOD	COMMENTS
1	x		768 kb/s		CIF	H.263	Full screen
2	x		1.5 Mb/s		CIF	H.263	Full screen
3	x		2 Mb/s	@	$\frac{3}{4}$	mp@ml	This is a horizontal resolution reduction only
4	x		2 Mb/s		$\frac{3}{4}$	sp@ml (**)	
5	x		TBD by ILSC			mp@ml	with errors TBD (see note 4)
6	x		TBD by ILSC			422p@ml	I only, with errors TBD (perhaps a lower bit rate)(see note 4)
7	x		4.5 Mb/s	@		mp@ml	
8 <sup>(*)</sup>	x	x	3 Mbit/s			mp@ml	
9 <sup>(*)</sup>	x	x	4.5 Mb/s			mp@ml	Composite NTSC and/or PAL
10		x	6 Mb/s	@		mp@ml	
11		x	8 Mb/s			mp@ml	Composite NTSC and/or PAL
12		x	8 & 4.5 Mb/s	@		mp@ml	Two codecs concatenated
13		x	12 Mb/s	@		mp@ml	
14		x	19/PAL(NTSC)-19/PAL(NTSC)-12 Mbit/s			422p@ml	PAL or NTSC 3 generations
15		x	50-50-....-50 Mbit/s	@		422p@ml	7th generation with shift / I frame
16		x	19-19-12 Mbit/s			422p@ml	3rd generations
17		x	n/a			n/a	Multi-generation Betacam with drop-out(4 or 5, composite/component)

Notes:

- 1) A, B = two different test groups, n/a = not applicable, TBD = to be determined
- 2) Different codecs should be used for the various HRCs as much as possible
- 3) All scenes are to be shown full size, specifically refers to HRC 1 and 2
- 4) For HRCs 5 and 6, artifacts are to be kept within the same quality range as the other impairments in the test. However multiple perceptible errors should occur within the sequence. Due to the technical concern over inclusion or not of these HRCs they are considered to be a pilot study. Results from these HRCs will be analyzed for consistency of its subjective data in order to determine if they should be included in the overall report of results.

(\*\*)If we don't find the SP it's enough to avoid the B frames.

(\*) Condition tested in both tests (low and high range of quality)



5) 525 and 625 versions are required for all HRCs.

If not enough resources are available for encoding, HRCs will be removed based on the following:

1. They are considered not in the 'core'.
2. They could represent an overlap with other conditions.
3. No volunteers are available for that particular HRC.

The ILSC will make final decisions on the HRC list.

## List of participants

Jamal Baina	TDF-C2R, France	Baina@c2rsmtmp.c2r.tdf.fr
Vittorio Baroncini	FUB, Italy	Vittorio@fub.it
John Beerends	KPN Research, Netherlands	j.g.beerends@research.kpn.com
Michael H. Brill	Sarnoff Corporation	mbrill@sarnoff.com
Laura Contin	CSELT, Italy	laura.contin@cselt.stet.it
Philip Corriveau	CRC, Canada	philc@dgbt.doc.ca
Frank de Caluwe	KPN Research, Netherlands	
Evain, Jean-Pierre	EBU	evain@ebu.ch
Charles Fenimore	NIST, USA	fenimore@eeel.nist.gov
David K. Fibush	Tektronix, USA	davef@tv.tv.tek.com
Alan S. Godber	Engineering Consultant, USA	agodber@idt.net
David Harrison	ITC, UK	harrison@itc.co.uk
John Libert	NIST, USA	libert@eeel.nist.gov
Jeffrey Lubin	Sarnoff Corporation, USA	jlubin@sarnoff.com
Al Morton	AT&T Laboratories, USA	acmorton@att.com
Ricardo M. Nishihara	CPQD Telebras, Brazil	nishihar@cpqd.com.br
Stephane Pefferkorn	CCETT (CNET), France	stephane.pefferkorn@cnet.francetelecom.fr
Mihir Ravel	Tektronix, USA	mihir.ravel@tek.com
Ann Marie Rohaly	Tektronix, USA	ann.m.rohaly@tek.com
Alexander Schertz	IRT, FRG	schertz@irt.de
Ernest Schmidt	Delta Information Systems, USA	eschmidt@delta-info.com
Massimo Visca	RAI, Italy	m.visca@rai.it
Andrew B. Watson	NASA, USA	abwatson@mail.arc.nasa.gov
Arthur Webster	NTIA/ITS, USA	awebster@its.blrdoc.gov
Stefan Winkler	EPFL, Switzerland	winkler@ltssg3.epfl.ch