

DRAFT MEETING RECORD
Video Compression Measurements Subcommittee G-2.1.6
Audio Video Techniques Committee G-2.1

Broadcast Technology Society
Institute of Electrical and Electronics Engineers
Sixteenth Meeting

NTIA/ITS U.S. Department of Commerce
325 Broadway
Boulder, CO 80303
April 24-25, 2000

Item 1 - Welcome and Introduction by Interim Chairman, of IEEE G-2.1.6.

Interim Chairman Alan Godber called the meeting to order at 3:25 PM. Attendees introduced themselves. See Appendix B for a list of attendees. Andrew Watson participated by telephone.

Item 2 - Approval of Draft Agenda

A contribution from Stephen Wolf and Andrew Watson was added to the agenda as Item 6.2. A liaison from ITU JWP-10/11Q provided by David Fibush was added to Item 5.0.

Item 3 - Review and Approval of Minutes of the Previous Meeting #13, July 27th, 1999

The Minutes from Meeting 15 were accepted as prepared.

Item 4 - Matters Arising from the Minutes

There were no matters arising from the minutes.

Item 5 - Update Report of ITU Video Quality Experts Group (VQEG) re results of the tests conducted, and results of the Ottawa meeting. - Arthur Webster, David Fibush, Phil Corriveau, Al Morton and other participants.

Ann Marie Rohaly reported that the VQEG Final Report was ratified, closing out the first phase of testing. An ad hoc group looked at designing a more discriminating full-reference video quality test. Other groups looked at reduced reference and no-reference quality tests for both broadcast and multimedia applications. A complete list of the Ad Hoc Committees, including chairs and committee email reflectors, is available in the [VQEG March 2000 Meeting Report](#) and the *Liaison Statement*, VQEG, IEEE Doc. G-2.1.6-6/110.

It was reported that it was difficult to find expertise in VQEG to develop a reduced-reference/no-reference multimedia quality measurement test plan. A questionnaire will be sent out. More progress was made on the reduced-reference/no-reference TV quality measurement testing. Work continues on the email reflector.

At the VQEG meeting there was a proposal by Andrew Watson to develop a "Performance-based Standard" that would be used to measure the performance of different models using a new standard data set.

Under the Phase 2 full-reference TV quality test plan, two viewing distances would be used. Some HRCs, such as multi-generation Betacam, H263, transmission errors and analog may be eliminated from the testing. It was noted that when these were eliminated from the original data set, the test results became worse, not better. There was another comment that the quality of the HRC scenes had to be looked at.

There was some concern that by splitting the efforts into four groups – TV quality full-reference, multimedia quality full-reference, TV quality reduced/no-reference and multimedia quality reduced/no-reference – the VQEG effort may be diluted. It was countered, however, that this was done on purpose to accelerate progress. By looking at all the options, the group can focus on which ones, if not all, it wants to work on.

Since the meeting, there is not been much activity on the VQEG email reflector. The next possible meeting date is in September, following the ITU meetings.

David Fibush presented the [*Liaison: Report on Activities of ITU-R JWP 10-11Q*](#), Tektronix, April 14, 2000, T1A1.1 /2000-0150, IEEE Doc. G-2.1.6/111. Topics covered included: results of the VQEG validation tests and future directions; use of a reference model for specification of objective picture quality measurements; coordinated action by ITU-T and ITU-R study groups; revision of draft Report on Objective Quality Evaluation; draft revision of ITU-R BT.500; rescaling of subjective scores; subjective listening test of intermediate audio quality; new test materials for SD and HD 60 Hz systems; and revision of ITU-R.BT.1210.

5.1 Further Discussion and Recommendations from the Subcommittee to the next meeting of VQEG

The committee was asked to provide comments on the relative importance of full reference, reduced reference and no reference video quality measurement systems. It was noted that reduced reference systems are being used as a monitoring scheme at the end of the transmission path, more for system monitoring than picture quality monitoring. DVB has defined a PID for test purposes where reduced reference information can be transmitted. There was a comment that this could perform the same purpose in monitoring compressed digital video paths as Vertical Interval Test Signals (VITS) perform in monitoring analog video paths. The general feeling of the committee was that picture quality measurement work with reduced references is useful. It was added that the ITU Study Group 9 effort split into three drafts – zero, reduced and full reference measurement methods.

Item 6 - Report of Task Force on "Defining A Unit of Measure & a Means of Calibration for Video Impairment", Chair, John Libert/Leon Stanger.

It was reported that John Libert's group at NIST working on measurement of video impairments has been disbanded and reassigned to other projects. Ann Marie Rohaly said that John Libert indicated he would continue to help VQEG on his own. Alan Godber said he had spoken to Charles Fenimore at NIST about whether writing a letter concerning John Libert's work with IEEE might lead NIST to change its mind on this. It was also suggested that the committee send a letter to John Libert thanking him for the work he has done with this group and with VQEG and invite him to participate as a private citizen. Arthur Webster suggested copies of the letters be sent to Peter Roitman at NIST.

ACTION ITEM: Alan Godber suggested sending a letter to John Libert thanking him for his work with the committee. Leon Stanger will look into this and make recommendations.

Andrew Watson presented his *Proposal: Measurement of a JND Scale for Video Quality*, IEEE Doc. G-2.1.6/112, April 24, 2000. This document is available on the IEEE G-2.1.6 web site. The presentation described two pilot experiments using a limited subset of the VQEG data.

After the presentation, the training of the viewers used in the testing was discussed. It was suggested that it is useful to give every observer a little bit of training before the testing begins. This could be accomplished by presenting the maximum strength stimulus to the viewer on the first trial. It would also give the viewer a good feeling that they can do the test.

There was concern that it won't be possible to get five thresholds out of some of the VQEG HRCs. It was mentioned, however, that while it was possible to get 8 JNDs out of HRC 15, it might only be possible to get one JND out of HRC 7. In this case, one JND was considered to be proportional to about 5 DMOS points.

The use of the data collected in the testing was discussed. Andrew Watson said the proposal did not state what would be done with the results. Unlike VQEG, there is no objective test plan or any method for comparing it to models. What we will have is data that describes the number of JNDs in the various HRC/SRC combinations. The logic of the experiment would imply that these numbers are absolute numbers than should be repeatable from lab to lab, observer to observer. As such, this data could be used to calibrate the VQEG data in terms of JND and could be compared with the objective scores in the VQEG tests. It would also enable models not measuring in JND to calibrate to JND.

It was questioned whether we will get absolute number from the tests. Andrew Watson replied that he expects the differences will be smaller than in DSCQS. Another question was how he would know the viewers used for the testing would be typical. The response was that there is a difference between an absolute scale and an accurate scale. JND are dimensionless. We will, however, know the degree of variability. Two observers on the first HRC/SRC combination were remarkably close. Observers, however, could be tested for basic visual acuity and color sensitivity. Another way to reduce variability would be to limit viewers to ages between 20 and 25 with 20/20 vision. It was recommended that these limits be documented so that another lab would be able to get similar results. Andrew Watson said the tests would follow VQEG procedures. It was noted that in the VQEG testing most labs had correlation above 0.9. One lab was much worse, but they were unable to find a reason for the discrepancy.

The *Proposal* listed four labs. Andrew Watson pointed out that they were listed not to indicate they would be available for the testing, but that they seemed interested in the work and had the equipment to do the experiment. It wasn't clear if the Tektronix lab would have the resources to collect the data, but it looked like the NTIA lab would be able to participate and funding may be available to use the Sarnoff Research Center lab. NASA has the facilities to do the work and Andrew Watson indicated they were going to proceed with the research. While some suggested it would be better to have more than one lab do the work, there was caution that if we want to do more tests in the future it wouldn't be good to overwork the other labs now.

Alan Godber said the committee needed to know what the result would be. Would it be a report or a test tape?

ACTION ITEM: Andrew Watson said he would write a couple extra pages to cover this.

The meeting was adjourned at 5:17 PM to resume April 25th at 8:30 AM

Item 6.2 – Two Objective Video Quality Metrics, Stephen Wolf and Andrew Watson.

Steve Wolf presented a report on ITS / NTIA Department of Commerce work on video quality metrics. This work is covered in Appendix A of *Two Objective Video Quality Metrics*, Stephen Wolf, ITS/NTIA/DOC and Andrew Watson, NASA Ames Research Center, ITU Study Group 9 Question 22/9, IEEE Doc. G-2.1.6/114. Additional information is available in a paper presented at SPIE, available at <http://www.its.bldrdoc.gov/n3/video/pdf/spie99.pdf>.

The presentation focused on metrics based on spatial gradient parameters. Tests showed the method was a significant estimator of video quality.

The recommended spatial-temporal (S-T) size is 8 horizontal pixels by 8 lines by 6 frames. The method has been found to be useful with up to 100,000:1 compression of feature extraction data. Tests found there was no need to go below 50:1 or 100:1. The difference was 0.01 for up to 300:1 to 400:1 compression.

Temporal collapsing is used to produce an objective quality parameter for a clip between 5 and 10 seconds in length. More research is needed to handle quality changes that occur during these 10 seconds.

Tests were done with eleven data sets as described in Appendix A. Steve Wolf said that he doesn't think correlation is a good way to check accuracy, preferring to use mean square error instead. Individual viewer's accuracy was about 1/5 of the quality scale. RMS error was about 1/25 of the quality scale, based on using 30 subjects.

In response to concerns about the accuracy spread, Steve Wolf noted that the VQEG data set was one small dataset of the total plot of points, probably one half of the scale. Another explained that the correlations were better than those of VQEG, but if you look at one portion of the plot/data set the correlations are not as good. The accuracy comes from stringing together the data sets. You can't make use of these correlations without knowing how the datasets were strung together. It isn't clear how you could do this, since the VQEG data occupies only the higher quality portion of the graph.

There was some agreement that these results can't be compared directly with VQEG. However, it was noted that there are formulas that can be used to make a direct comparison. Not only the range of data but the number of data points in the set must be considered. Steve Wolf replied should be done using the mean squared error. Chi-squared error was also suggested.

Steve Wolf proposed that HRC variance be analyzed as part of the testing. He also requested root-mean-square error be used instead of correlation. It was argued, however, that correlation should be included, if only for historical reasons.

Steve Wolf said he planned to post research results to the (VQEG?) reflector. He said that while he was not pushing this technique for in service measurements, it could be extended to that. The technique is being submitted to ITU Study Group 9. Additional information is available on the I.T.S. Video Quality Research home page at <http://www.its.bldrdoc.gov/n3/video/>.

Appendix B of the report covered the Digital Video Quality Metric, which Andrew Watson presented to this committee last year.

Item 7 - Further Discussion of Compression Measurement Methodologies.

Item 7.1 – Discussion of Future Work, Additional Assignments

The items covered this meeting were reviewed.

Andrew Watson will send the committee extra pages describing the tests. This committee can help by examining his proposal off-line and submitting comments and suggested modifications to him. Steve Wolf was interested, but can't commit until June or July. If he can get a copy of the software he may be able to run some viewers through some of the combinations.

Andrew Watson said he would push ahead without any guarantee of other labs participating. However, he cautioned that if people feel there should be changes he won't be able to go back to do it a second time. Any input should be sent by early June, if possible.

The procedures for giving viewers feedback were discussed. It was pointed out that the threshold might be different with feedback compared with when there is no feedback. It was asked if there was a procedure that would not give the viewer feedback. Andrew Watson replied that there are auditing procedures that could be used, but the purpose of the feedback is that it makes the viewers feel better and, in addition, there is the theoretical possibility that the compressed sequence could look better than the original sequence. Feedback is useful in identifying which is the test sequence. He said that we should use feedback for the first reason, but it wouldn't make a big difference in the results. While tests in psychology are usually double blind, in psychophysical tests you do give feedback. Others supported the use of feedback and it was suggested sacrificing some trials, four were suggested, for training.

It was recommended to have a post-test interview with the participants. This was agreed on, provided there was time. There was also a request to post the data on the reflector as an Excel spreadsheet.

Item 8 - Any Other Business.

This item was discussed prior to the resumption of discussion under Item 6 after the meeting resumed April 25.

An email from Herbert Bennett, regarding liaison to G-2.1.6 was presented (IEEE Doc. G-2.1.6/113). The email asked whether a liaison between the IEEE EDS and IEEE BTS would be worthwhile. From the discussion, it was decided to offer to make him the official liaison with between the IEEE EDS and the IEEE BTS.

ACTION ITEM: Doug Lung will send him information on G-2.1.6 with copies to Alan Godber and Arthur Webster.

Item 9 - Date(s) of Future Meeting(s).

The next meeting will be on Monday, the week of the T1A1 meetings in Schaumburg, Illinois. (Later determined to be August 7, 2000.)

The meeting was adjourned at 10:21 AM, April 25.

Submitted by:
H. Douglas Lung
Secretary

APPENDIX "A"

List of Documents Distributed

24-25 April 2000

Draft Agenda - IEEE Compression and Processing Subcommittee G-2.1.6, Sixteenth Meeting, Sunday-Monday, April 24-25, 2000, Alan Godber, Chairman, ([216m16an.pdf](#))

Draft Meeting Record, G-2.1.6, Compression and Processing Subcommittee, Fifteenth Meeting, January 23, 2000, Sunnyvale, CA, Doug Lung, Secretary, [IEEE Doc. G-2.1.6/109](#), April 17, 2000.

Liaison Statement, VQEG, April 18, 2000, IEEE Doc. G-2.1.6/110, April 24, 2000.

Liaison: Report on Activities of ITU-R JWP 10-11Q, Tektronix, April 14, 2000, T1A1.1 Doc. # [T1A1.1/2000-015](#), IEEE Doc. G-2.1.6/111, April 24, 2000.

Proposal: Measurement of a JND Scale for Video Quality, Andrew B. Watson, NASA Ames Research Center, IEEE Doc. G-2.1.6/112, April 24, 2000. Available at <http://vision.arc.nasa.gov/jnd/>

Email from Herbert S. Bennett regarding liaison between IEEE EDS and IEEE BTS, April 20, 2000, IEEE Doc. G-2.1.6/113, April 25, 2000.

Two Objective Video Quality Metrics, Stephen Wolf, ITS/NTIA/DOC and Andrew B. Watson, NASA Ames Research Center, ITU Study Group 9 Question 22/9, T1A1.1 Doc. IEEE Doc.# [T1A1.1/2000-021](#) G-2.1.6/114, April 25, 2000.

APPENDIX "B"
ATTENDANCE RECORD
24-25 April 2000

Name	Affiliation	Telephone	Fax	E-mail
Chairman: Alan Godber	Consultant	(732) 846-4476	(732) 846-4476	agodber@idt.net
Secretary: Doug Lung	Telemundo	(305) 884-9664	(914) 412-2886	dlung@transmitter.com
Dick Bobilin	Creative Communications	(732) 842-6250	(732) 220-8219	Bobilinccc@aol.com
David Fibush	Tektronix	(503) 628-3040	(503) 627-4486	davef@exgate.tek.com
Michel Poulin	Leitch	(416) 443-2716	(416) 445-4762	michel.poulin@leitch.com
Ann Marie Rohaly	Tektronix	(503) 617-3048	(503) 627-5177	ann.marie.rohaly@tek.com
Leon Stanger	DirecTV	(310) 726-4676		LStanger@compuserve.com
Arthur Webster	NTIA / ITS	(303) 497-3567	(303) 497-5969	webster@its.bldrdoc.gov
Stephen Wolf	NTIA / ITS	(303) 497-3771	(303) 497-5323	steve@its.bldrdoc.gov
Andrew Watson	NASA	(650) 604-5419	(650) 604-0255	abwatson@mail.arc.nasa.gov