

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
Fourteenth Meeting

The Westin Hotel
400 Corporate Drive, Radice Corporate Park
Fort Lauderdale, FL 33334-3642
November 1, 1999

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

The meeting was called to order at 1:40 PM by Interim Chairman Alan Godber.

Item 2 - Approval of Draft Agenda

The agenda was approved.

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

The [Minutes from Meeting 13](#) were approved as an accurate record of the last meeting.

Item 4 - Matters Arising from the Minutes

There was no discussion under this item.

Item 5 - Report of ITU Video Quality Experts Group (VQEG) re the tests conducted - Arthur Webster, David Fibush, John Libert, Al Morton and other participants.

David Fibush reviewed the *Preliminary Report from VQEG on the Validation of Objective Models of Video Quality Assessment*, Rapporteur of ITU-T Q11/12 and VQEG, 10 September 1999, IEEE Doc G-2.1.6/102, November 1, 1999.

The *Preliminary Report* showed that PSNR did a better job with high quality (studio quality) video than any of the proponents. It was not clear, however, that this is the final conclusion. Participants recognized items they could fix to improve the tests. There is more to do be done. The results demonstrate the need for what we (IEEE

G-2.1.6) are trying to do here -- provide a method for calibration. It was noted that with the wide variation in subjective test scores, it is questionable if anything can be compared to them.

David Fibush explained that VQEG decided the data from the tests should be available to anyone to use as they wish. There was some concern, however, about non-VQEG members using the data to create a system that would beat the test proponents. Once the results were seen and six of the proponents were close, there wasn't as much of a reason to keep it secret. None of the proponents performed as well as their pre-VQEG data would indicate.

The test results and methods of analyzing them were discussed, with the repeated caution that much work remained to be done. The plots on page 55 were done using curve fits on optimized data. John Libert said he would produce a final set of graphs, which will be submitted to the management of VQEG.

Other ITU groups, at the request of VQEG, have provided VQEG with guidance on future work. This could include tests with different bit rates, better curve fits, etc. David Fibush encouraged people to offer their ideas. If others are interested, they will support it and help supply resources. There was interest in broadcast applications as well as single-ended measurements. Many people, including Study Group 12 and even broadcasters, are interested in work at lower bit rates.

A copy of the liaison from Study Group 9 (SG9) to VQEG, *Liaison Report: User Requirements of an Objective Video Quality Measurement System for Cable Television; Draft New Recommendation J.ovq-req*, VQEG, 30 September 1999, IEEE Doc. G-2.1.6/103, November 1, 1999, was distributed. The document reports "SG9 has begun work on three draft new Recommendations covering objective quality methods with: 1) full reference material (J.fullref); 2) reduced reference material available (J.redref); and 3) no reference material available (J.noref). SG9 requested input from VQEG on these recommendations, particularly the full reference method, by the end of December 1999.

SG9 cancelled its interim meeting in January 2000. The next meeting will be in May. There was a comment that this time should be used to fine-tune the measurements and to determine how different methods work in different environments. If subjects don't see a difference between images, we shouldn't expect the objective measurements to see a difference. However, where subjects see a difference, the objective measurements should also show a difference.

Copies of *Call for Coordinated Action by Three Study Groups*, David Fibush, Tektronix, October 22, 1999, IEEE Doc. G.1.6/104, November 1, 1999 were distributed.

David Fibush explained that organizations were considering objective quality measurement standards ranging from material with significant errors and dropped frames to broadcast quality applications. Within these ranges, there are three different measurement methods, as outlined in the SG9 liaison report. The *Call for Coordinated Action by Three Study Groups* pointed out the need for all of the study groups to agree on what technologies they are going to use in each area. The contribution stated, "It would be a major disservice to those users if a different method(s) were recommended by each of the three study groups."

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

The concept of three different measurement methodologies and two different quality ranges was discussed. It was argued that the two quality ranges are really a continuum, as is the measurement methodology, where the bandwidth of the reference signal is the continuum.

There was also a discussion concerning the selection of scenes for the low-quality and high-quality test ranges in the VQEG test. For example, in the low quality set, HRC10 was better than any other scene in the set. In the high quality set, the multi-generation Betacam sequence had lower scores than most others in the set and was not what would be considered broadcast quality video. Some of the proponents' results were upset by one or two points. It was suggested it might be interesting to look at this and perhaps go back and reclassify the quality of some scenes.

The discussion turned to future work. The final report should be done in December. Once people review these results, there may be more tests. It was noted that there is interest in looking at the data in different ways to give more information in the final report. The idea is to analyze the existing data, not generate new data. This group (G-2.1.6) should indicate what further testing should be done in what areas.

Since one method did not emerge as better, there was a question about what will happen now. In discussion, there was a comment that some proponents might combine their methods. Doing threshold measurements on the VQEG data was also discussed. While John Libert was finding it difficult to generate critical mass with work on JND, Dr. Watson had a lot of interest in the threshold work. If the visibility threshold is well defined for any sequence, we will have to look at why the subjective data from the VQEG test isn't well defined.

In summarizing the reaction to the VQEG process, it was explained that everyone was a bit disappointed. However, it was also pointed out that no one had anywhere near this volume of data to look at before. The tests served a function, but we are not done yet. Everyone can improve their model for the next set of tests. Correlation between the subjective labs is the target for the objective models.

One idea mentioned was to stretch or compress individual subject's scales to match a mean scale. This would be useful because individual subjects make use of different portions of the 100-unit scale. People may not judge the top and bottom of the scale the same. If a subject uses only half the scale, the differences reported will be less. This scaling would be done only among subjects that have seen the same material. There was a comment that one lab used a completely different scaling method than the others - knobs rather than sliders. Another noted that if the result didn't change from test to test, some labs counted it as a vote with the same scale while others counted it as a missing vote. Some method is needed to show a decision was made.

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

Leon Stanger was not present. The results of the last G-2.1.6 meeting were recapped, leading to the new plan proposed by Andrew Watson, Ann Marie Rohaly and John Libert. The new plan had not been completed by this meeting, but John Libert agreed to talk about the basic idea behind it.

Under the modified plan, we would produce impaired video exhibiting various degradations. Looking at some number of samples at different impairment levels, we ask the subject to provide a numerical rating between the samples at impairment levels. With these numerical ratings and statistical analysis, we can determine where, if anywhere, the JND might be along the scale. At the end of the testing, we want a function that creates a continuous curve that describes what delta impairment is needed to see a difference between two levels of impairment. The results will allow us to generate calibrated test material.

John Libert said that by January 2000, his group, including Andrew Watson, Ann Marie Rohaly and Phil Corriveau, should have the methodology scoped out. He suggested we wait for it before commenting.

6.1 Further discussion and action

There was concern about the concept of making judgements over wide ranges of impairment rather than focusing on JNDs (Just Noticeable Differences).

John Libert responded that group JND measurements are not practical. You can't modify the stimuli in advance. It might be at a JND level. You have to do it one observer at a time, where you have an adaptive scheme so that you don't waste time on trials outside the threshold. This would involve too many trials to be practical. He pointed out that is not scaling. You are collecting more information from the subject. What you are looking for is a statistical distribution to allow you to infer where in the interval there is a JND. Each of the estimates points toward a region. Within the interval, you produce samples at whatever density you want, take every pair, and get a rating.

The problem with looking at a few JNDs is that you can't draw a reasonable function between three or four points. John Libert felt this method would be more reliable than ITU Rec.500 studies. He explained that the problem with most subjective measurements is that they ask for an absolute magnitude. In this case, we are looking for the magnitude of difference. We will have more control over the quantity we are varying through a mix of the various impairments.

Another comment was that different people might judge different impairments differently. The solution to this would be to synthesize scenes with specific impairments. John Libert answered that one problem with synthesis is getting someone to generate the scenes. He asked whether it is better to do something that is an improvement

on the VQEG test or do something completely different. Since we are allowing a preliminary test, we might as well do VQEG first.

Leon Stanger will be asked for input on the modified plan and will be list as a co-author on the paper covering it.

Item 7 - Report of Task Force on "Selecting Test Material and Test Labs for a Unit of Measure and a Means of Calibration for Video Impairment", Chair, John Libert.

John Libert reported on his paper, *Simulation of Graded Video Impairment by Weighted Summation: Validation of the Methodology*, preprint, SPIE Conference on Multimedia Systems and Applications II, 20-22 September 1999, John M. Libert, Charles P. Fenimore, and Peter Roitman, National Institute of Standards and Technology, IEEE Doc. G.1.6/105, November 1, 1999.

The paper describes methods for producing impaired video. One method varies the bit-rate of an MPEG encoder to generate impairments. PSNR was used to come up with a numerical measure of impairment at a given bit rate. Given measurements of a sampling of bit rates, from minimum to maximum, it is possible to describe a function by which you can predict the bit-rate parameter required to get an output sequence with a target impairment. This worked for a variety of sequences.

7.1 Further discussion and action

John Libert was questioned about the problem of multiple definitions for PSNR. How was color handled? He responded that the image has twice as many luminance points as chroma points, so this weighted the result. In this case, however, it doesn't matter since all that is being judged is a quantitative value of distortion.

There was also concern that at lower bit rates, people might see a change in quality with a very small change in bit rate. John Libert answered that he was looking for a discontinuity, but didn't find one, at least using the Test Model 5 MPEG encoder. Nothing would prevent you from having different quality curves for different portions of the bit rate curve. It would be interesting to obtain more data points using a different encoder.

Item 8 - Further Discussion of Compression Measurement Methodologies.

There was no discussion under this item.

Item 9 - Any Other Business.

There was no other business.

Item 10 - Date(s) of Future Meeting(s).

Because some important members of the committee will be attending and presenting papers at the Photonics West conference in San Jose, California, it was suggested we consider holding our next meeting before the conference. Saturday, January 22 or Sunday, January 23, in San Jose, was recommended as a possible meeting date. John Libert will talk to Andrew Watson and Ann Marie Rohaly about attending our meeting. The T1A1 meetings are scheduled from January 24 to 28 and conflict with the Photonics West conference. There was a motion to adjourn, which was seconded. The meeting was adjourned at 5:25 PM.

Submitted by:
H. Douglas Lung
Secretary

APPENDIX "A"

List of Documents Distributed

1 November 1999

Draft Agenda - IEEE Compression and Processing Subcommittee G-2.1.6, Fourteenth Meeting, Monday, November 1, 1999, Alan Godber, Chairman. ([216m14an.html](#))

Draft Meeting Record, G-2.1.6, Compression and Processing Subcommittee, Thirteenth Meeting, July 27, 1999, Minneapolis, MN, Doug Lung, Secretary, [IEEE Doc. G-2.1.6/101](#), November 1, 1999.

Preliminary Report from VQEG on the Validation of Objective Models of Video Quality Assessment, Rapporteur of ITU-T Q11/12 and VQEG, 10 September 1999, T1 Document [T1A1.5/99-103](#), IEEE Doc. G-2.1.6/102, November 1, 1999.

Liaison Report: User Requirements of an Objective Video Quality Measurement System for Cable Television; Draft New Recommendation J.ovq-req, VQEG, 30 September 1999, T1 Document [T1A1.5/99-104](#), IEEE Doc. G-2.1.6/103, November 1, 1999.

Call for Coordinated Action by Three Study Groups, David Fibush, Tektronix, October 22, 1999, T1 Document [T1A1.5/99/105](#), IEEE Doc. G.1.6/104, November 1, 1999.

Simulation of Graded Video Impairment by Weighted Summation: Validation of the Methodology, preprint, SPIE Conference on Multimedia Systems and Applications II, 20-22 September 1999, John M. Libert, Charles P. Fenimore, and Peter Roitman, National Institute of Standards and Technology, [IEEE Doc. G.1.6/105](#), November 1, 1999.

APPENDIX "B"
ATTENDANCE RECORD
1 November 1999

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		dlung@transmitter.com
David Fibush	Tektronix	(503) 628-3040	(503) 627-4486	davef@exgate.tek.com
John Libert	NIST	(301) 975-3828		john.libert@nist.gov
Al Morton	AT&T Labs	(732) 420-1571	Call for number	acmorton@att.com
Wallace Murray	Ameritech	(313) 983-8421	(313) 983-8649	wallace.w.murray@ameritech.com
Michel Poulin	Leitch	(416) 443-2716	(416) 445-4762	michel.poulin@leitch.com
Rick Redford	ABC	(212) 456-4450		rick.redford@juno.com
Ernest Schmidt	Delta Information Sys.	(215) 657-5270 x166	(215) 657-5273	eschmidt@delta-info.com
Dick Streeter	Consultant/CBS	(908) 791-9876	(908) 791-4878	rstreeter@att.net