UIT - Secteur de la normalisation des télécommunications
ITU - Telecommunication Standardization Sector
UIT - Sector de Normalización de las Telecomunicaciones

Commission d'études ;Study Group;Comisión de Estudio} 12

Contribution tardive;Delayed Contribution ;Contribución tardía} **999**

Texte disponible seulement en ;Text available only in;Texto disponible solamente en} English

Question: 11/12, 10/12

SOURCE[1]: Rapporteur

TITLE: Evaluation of new methods for objective testing of video quality: subjective test plan

Abstract

This contribution presents a subjective test plan that has been drafted by members of the ITU VQEG (Video Quality Experts Group) ad hoc committee for the subjective test plan. This test plan is offered to the participating ITU Study Groups (ITU-T Study Groups 9 and 12 and ITU-R Study Group 11) for further review and comment. The subjective test plan will be used to evaluate video quality in the bit rate range of 768 kbit/s to 36 Mbit/s. In conjunction with the objective test plan, it will be used to evaluate several proposed methods for objectively assessing video quality by measuring the correlation between subjective and objective assessments. It is expected that this test plan will be included in new Draft Recommendations in the area of video quality, probably as an annex.

[1] Contacts: Arthur Webster,
Rapporteur
Tel: +1 303 497 3567
Fax: +1 303 497 5323
E-mail: webster@its.bldrdoc.gov

Philip Corriveau
Editor
Tel: +1 613-998-7822
Fax: +1 613-998-7823
E-mail: philc@dgbt.doc.ca

VQEG SUBJECTIVE TEST PLAN

1 INTRODUCTION

A group of experts from three groups, the International Telecommunications Union (ITU-R) SG11, ITU-T SG9, and ITU-T SG12 assembled in Turin Italy on 14-16 October, 1997 to form the Video Quality Experts Group (VQEG). The goal of the meeting was to create a framework for the evaluation of new objective methods for video quality evaluation. Four groups were formed under the VQEG umbrella: Independent Labs and Selection Committee, Classes and Definitions, Objective Test Plan, and Subjective Test Plan. In order to assess the correlation between objective and subjective methods, a detailed subjective test plan has been drafted.

The purpose to subjective testing is to provide data on the quality of video sequences and to compare the results to the output of proposed objective measurement methods. This test plan provides a common criteria and process to ensure valid results from all participating facilities.

2 TEST MATERIALS

2.1 SELECTION OF TEST MATERIAL

The selection of sequences will be controlled and completed by the Independent Labs and Selection Committee (ILSC) in a time frame yet to be determined. Twenty source sequences and 20 Hypothetical Reference Circuits (HRC) are to be used in the testing. Following is a list of criteria for the selection of test material:

- at least one sequence must stress colour
- one still sequence
- several film sequences
- several sequences containing scene cuts
- several sequences containing motion energy and spatial detail
- at least one sequence containing text ***(What language will be used for the text or should it be numbers) - A concern raised by Al Morton.***

I would propose to use a very simple English text everybody understands like „My name is John“. I think that the assessors should read the text. This is probably not the case if the text is only a mixture of numbers.

- all Sources must be clean - use of noisy Sources is not permitted
- sequences must span the range of criticality and be representative of regular viewing material
- the introduction of transmission errors must not violate quality range - local errors can be bad but not unduly so

Based on the assumption that there will be 21 [24 to account for overlap between low and high ranges] HRC's and 24 source sequences, a total set of 504 processed sequences will be available. According to current experimental designs all 504 sequences will be tested.

2.2 SEGMENTATION OF TEST MATERIAL

Since there are two standard formats 525:60 and 625:50, the test material could be split 50/50 between them. Also, two bit rate ranges will be covered with two separate tests in order to avoid compression of subjective ratings. Therefore, the first test will be done using a low bit rate range of 768 kb/s - 8 Mb/s and 4.5 Mb/s (1-12 Table 1) with 12 processed sequences. A second test will be done using a high bit rate range of 4.5 Mb/s - 36 Mb/s (13-21 Table 1) plus 3 HRC's ***(6,9,10) or (6,9,12)*** that overlap with the low bit rate range with 12 processed sequences.

2.2.1 DISTRIBUTION OF TESTS OVER FACILITIES

Each test tape will be assigned a number so that we are able to track which facility conducts which test. It will be taken into account when making the test tapes that some processed sequences will be repeated within one test tape and between test tapes for control purposes. The tape number will be inserted directly into the data file so that the data is linked to one test tape.

2.3 HYPOTHETICAL REFERENCE CIRCUITS (HRC)

TABLE 1 LIST OF HRC'S (This table has not been finalized)

	BIT RATE	RES	METHOD	COMMENTS
1	768 kb/s	CIF	H.263	
2	1.5 Mb/s	CIF	H.263	
3	2 Mb/s	¾	mp@ml	This is a horizontal resolution reduction only
4	2 Mb/s	¾	sp@ml	
5	3 Mb/s		mp@ml	
6	4.5 Mb/s		mp@ml	
7	4.5 Mb/s		mp@ml	Composite NTSC and/or PAL
8	4.5 Mb/s	¾	sp@ml	At least one commercial encoder does this
9	6 Mb/s		mp@ml	
10	8 Mb/s		mp@ml	
11	8 Mb/s		mp@ml	Composite NTSC and/or PAL
12	8 & 4.5 Mb/s		mp@ml	Two codecs concatenated
13	12 Mb/s		mp@ml	
14	18 Mb/s		422p@ml	IB only, true SX encoder preferred
15	21 Mb/s		422p@ml	Large GOP, new EBU standard
16	36 Mb/s		422p@ml	I only
17	36 Mb/s		422p@ml	IB only, Eight generations, spatial and GOP shifts
18	n/a		n/a	VHS (must be fully time base corrected)
19	n/a		n/a	Multi-generation Betacam (4 or 5, composite/component)
20			mp@ml	with errors TBD (to be determined)
21			422p@ml	I only, with errors TBD (perhaps a lower bit rate)

For 20 and 21, artifacts are to be kept within the same quality range as the other impairments in the test.

2.3.1 PROCESSING AND EDITING SEQUENCES

The sequences required for testing will be produced based on the block diagram shown in Figure 1. Rec. 601 Source component will be converted to Composite (for HRC 7 & 11 only) and passed through different MPEG-2 encoders at the various HRC's with the processed sequences recorded on a VTR (potentially D1).

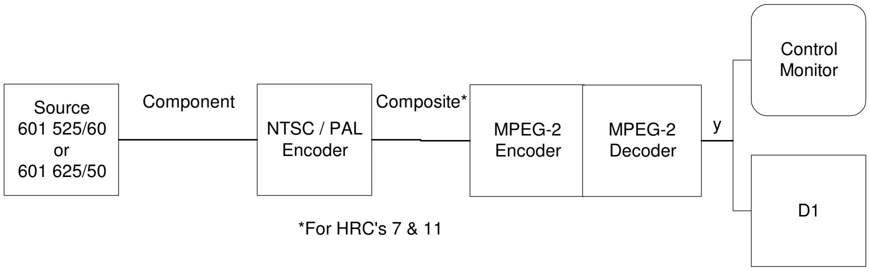


FIGURE 1 SEQUENCE PROCESSING

The processed sequences are then edited onto test tapes (potentially D1) using edit decision lists leading to the production of randomizations distributed to each test facility for use in subjective testing sessions.

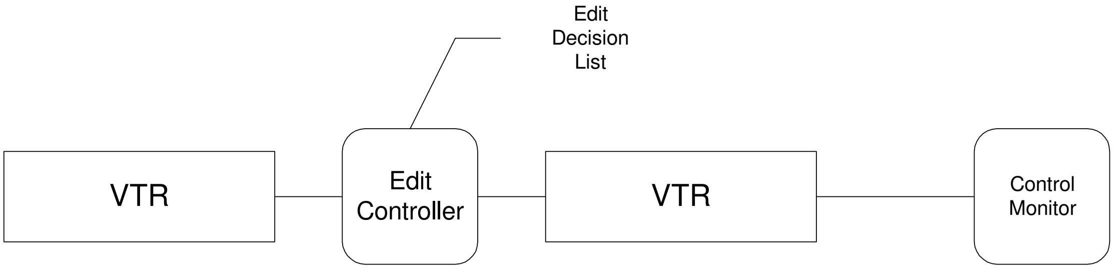


FIGURE 2 EDIT PROCESSING

2.3.2 RANDOMIZATIONS

For all test tapes produced, a detailed Edit Decision List should be created with an effort to:

- spread conditions and sequences evenly over tapes for any given session
- try to have a minimum of 2 trials between the same sequence
- have a maximum of 2 consecutive presentations: (S/P S/P; S/P S/P, P/S P/S,)
- have a maximum of 2 consecutive conditions, i.e. HRC's
- ensure that no sequence is preceded or followed by any other specific sequence more than once in order to minimize contextual effects

2.4 PRESENTATION STRUCTURE OF TEST MATERIAL

Due to fatigue issues, the sessions must be split into three sections: three 30 minute viewing periods with two 20 minute breaks in between. This will allow for maximum exposure and best use of any one viewer. A typical session would consist of:

5 warm-up trials + 28 test trials
20 minute break
3 reset trials + 30 test trials
20 minute break
3 reset trials + 30 test trials

This yields a group of up to 6 subjects evaluating 88 test trials at one time. The subjects will remain in the same seating position for all 3 viewing periods.

The individual test trials will be structured using the ABAB style shown in Figure 3:

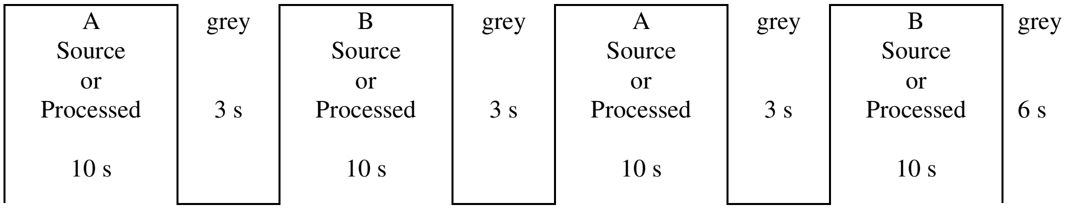


FIGURE 3 PRESENTATION STRUCTURE OF TEST MATERIAL

3 THE DOUBLE-STIMULUS CONTINUOUS QUALITY-SCALE METHOD

3.1 GENERAL DESCRIPTION

The Double Stimulus Continuous Quality Scale (DSCQS) Method presents two pictures (twice each) to the assessor, where one is a source sequence and the other is a processed sequence. See Figure 3. A source sequence is unimpaired whereas a processed sequence may or may not be impaired. The sequence presentations are randomized on the test tape to avoid the clustering of the same conditions or sequences. After the second presentation of the sequences, participants evaluate the picture quality of both sequences using a grading scale (DSCQS).

3.2 GRADING SCALE

The DSCQS consists of two identical 10 cm graphical scales which are divided into five equal intervals with the following adjectives from top to bottom: Excellent 100-80, Good 79-60, Fair 59-40, Poor 39-20 and Bad 19-0. (Note: adjectives will be written in the language of the country performing the tests.) The scales are positioned in pairs to facilitate the assessment of each sequence, i.e. both the source and processed sequence. The viewer records his/her assessment of the overall picture quality with the use of pen and paper provided. Figure 4, shown below, illustrates the DSCQS.

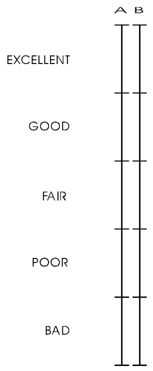


FIGURE 4 DSCQS (NOT TO SCALE)

4 VIEWING CONDITIONS

Viewing conditions should comply with those described in International Telecommunications Union Recommendation ITU-R BT.500-7. An example of a viewing room is shown in Figure 5. Specific viewing conditions for subjective assessments in a laboratory environment are:

– Ratio of luminance of inactive screen to peak luminance: ≤ 0.02
– Ratio of the luminance of the screen, when displaying only black level in a completely dark room, to that corresponding to peak white: ≅ 0.01

- Display brightness and contrast: set up via PLUGE (see Recommendations ITU-R BT.814 and ITU-R BT.815)
- Maximum observation angle relative to the normal*: 30^0
- Ratio of luminance of background behind picture monitor to peak luminance of picture: ≅ 0.15
- Chromaticity of background: D_{65}
- Other room illumination: low

This number applies to CRT displays, whereas the appropriate numbers for other displays are under study.

The monitor size selected to be used in the subjective assessments is a 19” Sony BVM 1910 or 1911 or any other 19” Professional Grade monitor.

The viewing distance of 5H selected by VQEG falls in the range of 4 to 6 H, i.e. four to six times the height of the picture tube, compliant with Recommendation ITU-R BT.500-7.

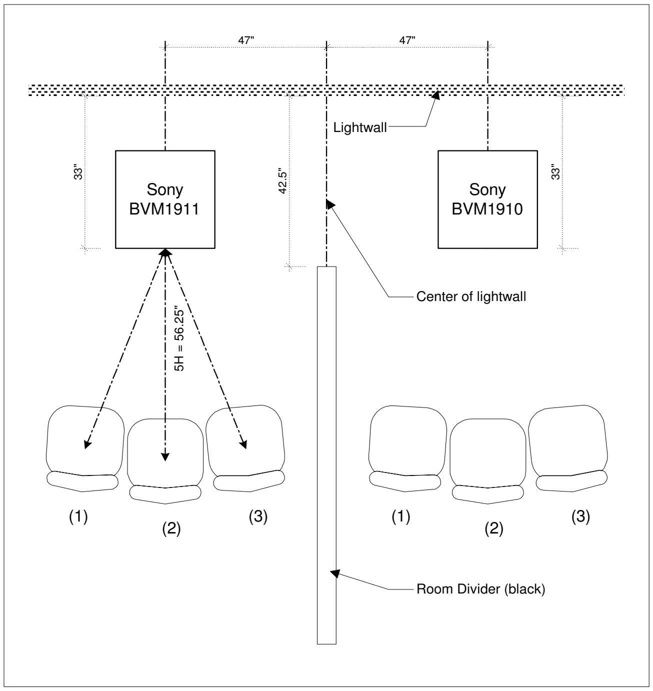


FIGURE 5 VIEWING ROOM AT THE CRC[*]

[*] As an example, this diagram shows the viewing room used for subjective tests at the Communications Research Centre (CRC).

4.1 INSTRUCTIONS TO VIEWERS FOR QUALITY TESTS

The following text could be the instructions given to subjects.

In this test, we ask you to evaluate the overall quality of the video material you see. We are interested in your opinion of the video quality of each scene. Please do not base your opinion on the content of the scene or the quality of the acting. Take into account the different aspects of the video quality and form your opinion based upon your total impression of the video quality.

Possible problems in quality include:

- poor, or inconsistent, reproduction of detail;
- poor reproduction of colours, brightness, or depth;
- poor reproduction of motion;
- imperfections, such as false patterns, or “snow”.

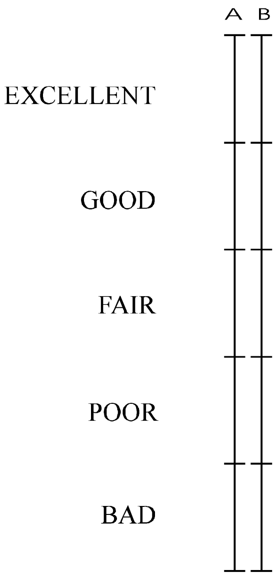
The test consists of a series of judgment trials, each consisting of four presentations of the same piece of video material.

Each trial will be announced verbally by number. The first presentation of a trial will be announced as “A”, and the second as “B”. This pair of presentations will then be repeated, thereby completing a single trial. Please note that one of the presentations, A *or* B, will be a Source picture that shows the video material in a high quality format, with the other being a Processed picture that shows the same video material in the format being tested.

We will now show you four demonstration trials.

DEMONSTRATION TRIALS PRESENTED AT THIS POINT

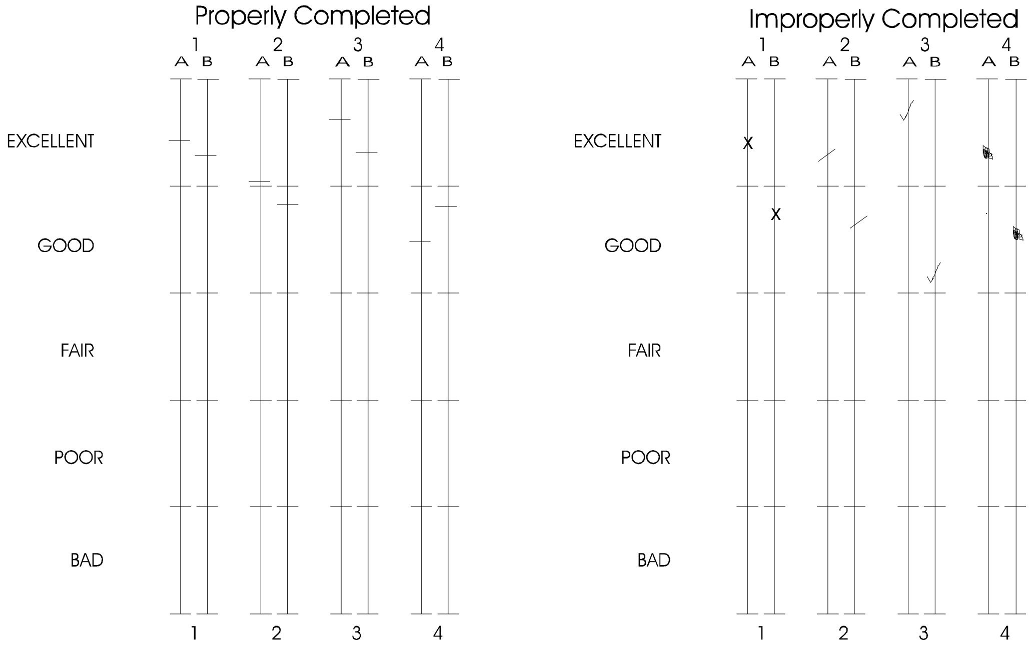
In judging the overall quality of the presentations, we ask you to use judgment scales like the samples shown below.



SAMPLE QUALITY SCALE

As you can see, there are two scales for each trial, one for the “A” presentation and one for the “B” presentation, since both the “A” and “B” presentations are to be judged.

The judgment scales are continuous vertical lines that are divided into five segments. As a guide, the adjectives "excellent", "good", "fair", "poor", and "bad" have been aligned with the five segments of the scales. You are asked to place a single horizontal line at the point on the scale that best corresponds to your judgment of the overall quality of the presentation (as shown in the example).



You may make your mark at any point on the scale which most precisely represents your judgment.

In making your judgments, we ask you to use the first pair of presentations in the trial to form an impression of the quality of each presentation, but to refrain from recording your judgments. You may then use the second pair of presentations to confirm your first impressions and to record your judgments in your Response Booklet.

5 VIEWERS

A minimum of 20-25 non-expert viewers should be used. The term non-expert is used in the sense that the viewers' work does not involve television picture quality and they are not experienced assessors. All viewers will be screened prior to participation for the following:

- normal (20/40) visual acuity or corrective glasses (per Snellen test or equivalent)
- normal contrast sensitivity (per Pelli-Robson test or equivalent)
- normal colour vision (per Ishihara test or equivalent)
- familiarity with the language sufficient to comprehend instruction and to provide valid responses using semantic judgment terms expressed in that language.

6 DATA

6.1 RAW DATA FORMAT

Depending on the facility conducting the evaluations, data entries may vary, however the structure of the resulting data should be consistent among laboratories. An ASCII format data file should be produced with certain header information followed by relevant data pertaining to the ratings/judgments ***including the results of the warm-up and reset trials*** see below:

In order to preserve the way in which data is captured, one file will be created with the following information:

RAW DATA

Subject Number[2]	SxHRCy		SxHRCy		SxHRCy	
	source	process	process	source	source	process
1001	95.1	62.3	71.5	20.4	75.8	49.3…
1002	88.6	60.4	75.1	21.2	77.0	51.3…
.						
.						

All scene and HRC combination will be identified in the first row of the file. All these files should have extensions ".dat". This file will include the test results for warm-up and reset trials. These also will be labeled. The files should be in ASCII format and/or Excel format.

6.2 SUBJECT DATA FORMAT

The purpose of this file is to contain all information pertaining to individual subjects who participate in the evaluating. The structure of the file would be the following:

Subject Number[1]	Tape Number	Month	Day	Year	Age	Gender*
1001	01	02	12	98	25	2
1002	01	02	12	98	32	1

*Gender where 1=Male, 2=Female

6.3 DE-RANDOMIZED DATA

In a normal situation for the statistical analysis of data it is nice to have the data set sorted in order of scene and HRC combination. It is proposed that if possible each lab produce a data file with sorted data to resemble the following:

				SORTED DATA POINTS		
Subject Number	Tape	Age	Gender	S1HRC1	S1HRC2	S1HRC3..
1001	01	27	2	78.0	53.5	49.1

7 DATA ANALYSIS

The data analysis for the subjective test results will include some or all of the following:

- Spearman's Correlation Coefficient
- Ranked Correlation Coefficient
- RMS error
- Weighted RMS Error
- Some other non parametric method
- Anova/Manova: Analysis of Variance - an inferential statistical technique used to compare differences between two or more groups with the purpose of making a decision that the independent variable influenced the dependent variable.

[2] The first digit of Subject Number will indicate the lab which conducted those evaluations

The following bold italic message is just a proposal made by Coleen and Philip and has not been accepted by VQEG as of yet:

- This test design tests all possible combinations of HRC and scene (in 625/50 or 525/60 video format.
- The number of HRC's in the low bit-rate test has been changed.
- The number of HRC's in the high bit-rate test has been changed.
- The division of HRC's between the low bit-rate test and the high bit-rate test has been changed.
- The division of HRC's between the two tests has limited the overlap between the two to only 3 HRCs ***These three overlapping HRC's need to be agreed upon***.
- The number of scenes has been changed.
- A 50/50 split of scenes between NTSC and PAL is assumed. ***If the ILSC can not accomplish this, the design will have to change***.
- It is assumed that HRC's 20 and 21 (transmission impairments) will be included.
- Number of subjects has been changed.
- Any changes in the number of scenes, HRC's, division of scenes between 625/50 and 525/60, or the low bit-rate/high bit-rate boundaries will necessitate a new test design.

Test Dimensions (or factors):

HRC: There are 21 HRC's as shown in Table 1. However, the HRC dimension is divided into two groups, the low bit-rate group, and the high bit-rate group. This immediately gives two individual subjective tests that need to be designed and conducted separately. Thus,

HRC 1-12:	Twelve low bit-rate test HRC's
HRC 13-21 + 3 TBD HRC's	Twelve high-bit rate test HRC's

This is a change from decisions made at the October VQEG meeting, but a balanced test design necessitates this division of HRC's.

SCENES: There are 24 scenes used in this subjective test. This is a change from decisions made at the October VQEG meeting, but a balanced test design necessitates that the number of scenes be divisible by 6 (2 video formats * 3 labs). As with the HRC dimension, the scene dimension is also divided into two groups, the 525/60 and 625/50 video formats. The test design assumes a 50/50 split of scenes between the two video formats. If the ILSC is not able to achieve this split, the balance of the design will be affected. The test design also assumes that each subset of scenes (525/60 and 625/50) will span a range of coding difficulty (perhaps quantified by something such as criticality).

SUBJECTS (within lab): Some testing conditions (scene-HRC pair) will be repeated among three labs (within one of the bit-rate tests), and some will be repeated among six labs (for conditions common to both bit-rate tests). Thus, the number of subjects is constrained to be divisible by 6. We thus require 24 subjects per condition. This will allow us to take sub-samples of subjects from these repeated conditions that have the same number of subjects as conditions that are not repeated. This sub-sample of subjects can then be evenly distributed from all labs. For example, a condition (c1) tested at only one lab will have 24 subjects. A condition (c2) repeated over six labs (this condition would be in the HRC's that overlap between the two bit-rate tests) will have 24*6 = 144 subjects. For balance in the ANOVA (for the HRC and scene factors), this condition needs only 24 subjects. To get this sub-sample of 24 subjects, 144-24 = 120, 120/6 = 20, so 4 subjects (24-20) would be used from each of the six labs maintaining balance between labs and subjects.

LABS: To determine any significant lab effects, at least three labs must test a set of common conditions so that comparison can be made between the three labs. Three labs would result in three comparisons, lab1 with lab2, lab1 with lab3, and lab2 with lab3. Thus, this test design calls for three labs for each video format. The three labs within a video format must be unique, but a given lab could be part of both video format lab pools.

HRC X Scene:
HRC X Lab
Scene X Lab
HRC X Subject (within lab)
Scene X Subject (within lab)
HRC X Scene X lab

Test Balance:
The ANOVA analysis is greatly simplified if the test design is balanced. This section discusses what is necessary to achieve this balance.

Definitions:

# comm. cond.: conditions common to all 3 labs testing a given bit-rate range with a given video format (525/60 or 626/50).

# lab cond.: all conditions unique to each of the 3 labs testing a given bit-rate range with a given video format (525/60 or 626/50). It is the difference between the total number of conditions and the number of common conditions.

Balance Constraints:

# comm. cond./# scenes = integer
# comm. cond./# HRC's = integer
*# lab cond./(3 * # scenes) = integer*
*# lab cond./(3 * # HRC's) = integer*

# lab cond./3 must be divisible by both the # scenes and the # of HRC's.
# comm. cond. Must be divisible by both the # scenes and the # of HRC's.
# lab cond./3 + # comm. cond. = # cond./subj. per lab.

Test Design:

There are four tests that need to be run:

Test1N – low bit rate (HRC 1-12) 525/60 (NTSC) format, the N is for NTSC, but it is meant to indicate 525/60 format (component or composite as HRC dictates).
Test1P – low bit rate 625/50 (PAL) format, the P is for PAL, but it is meant to indicate 625/50 format (component or composite as HRC dictates).
Test2N – high bit rate (HRC 13-21 + 3 TBD) 525/60 format
Test2P – high bit rate 625/50 format

If the HRCs are arranged by increasing quality, as they are in Table 1 of the test plan, and the scenes are arranged by decreasing coding difficulty and video format, we get the Matrix shown in Matrix 1. With this organization, within a given video format, the upper left-hand corner is lowest quality (lowest bit rate with most complex scene), and the lower right hand is highest quality (highest bit rate with least complex scene). The matrix shows a 50/50 split of scenes between the two video formats across all HRC's.

The following discusses the design of one low bit-rate test and one high bit-rate test. The design will be the same for both video formats.

Test 1 – low bit rates: The VQEG agreed that the low bit-rate test would span HRC's 1-11. However, for this test to achieve balance, the scene-HRC matrix must be 12X12. Thus, the HRC's have been changed to HRC 1-12. Also, the number of scenes has been increased from 10 (20/2) to 12 (24/2).

There are 144 conditions in this test, (12 scenes of a given video format and 12 HRC's).

144 HRC 1-12.
*9 conditions (3 scenes/HRC * 3 HRC's) over 3 TBD HRC's will be common to both the low bit-rate and high bit-rate tests.*

As per balance constraints discussed above, there are 36 common conditions (common among all three labs testing this bit-rate range).

36 cond./12 scenes = 3 HRCs/scene
36 cond./12 HRCs = 3 scenes/HRC
144 – 36 = 108, 108/3 = 36 unique conditions/lab
*36 cond./12 scenes = 3 HRCs/scene * 3 labs = 9 + 3 (common) = 12 HRCs/scene*
*36 cond./12 HRCs = 3 scenes/HRC * 3 labs = 9 + 3 (common) = 12 scenes/HRC*

144 – 36 = 108, 108/3 + 36 = 72 conditions/subject per lab

The distribution of conditions over labs is shown in Matrix 2.

Test 2 – high bit rates: The VQEG agreed that the high bit-rate test would span HRC's 6-21. However, for this test to achieve balance, the scene-HRC matrix must be either 12X12 or 12X24. The 12X12 matrix seemed the more reasonable approach. Thus, this test is defined at HRC's 13-21 plus 3 other HRC's that will overlap the low bit-rate test. HRC's 8, 9, and 10 are suggested as they cover three bit rates in the original area of overlap. However, the 3 overlapping HRC's can be chosen from any of the 12 HRC's in the low bit-rate test (HRC's 1-12). Also, the number of scenes has been increased from 10 (20/2) to 12 (24/2). This makes all four tests (two bit-rate ranges and two video formats) the same.

There are 144 conditions in this test, (12 scenes of a given video format and 12 HRC's).
144 HRC 13-21 + 3 TBD HRC's.
9 conditions over 3 TBD HRC's will be common to both the low bit-rate and high bit-rate tests.

As per balance constraints discussed above, there are 36 common conditions (common among all three labs testing this bit-rate range).

36 cond./12 scenes = 3 HRCs/scene
36 cond./12 HRCs = 3 scenes/HRC
144 – 36 = 108, 108/3 = 36 unique conditions/lab
36 cond./12 scenes = 3 HRCs/scene * 3 labs = 9 + 3 (common) = 12 HRCs/scene
36 cond./12 HRCs = 3 scenes/HRC * 3 labs = 9 + 3 (common) = 12 scenes/HRC

144 – 36 = 108, 108/3 + 36 = 72 conditions/subject per lab

The distribution of conditions over labs is shown in Matrix 2.

Conditions and Subjects: (for a given video format, for both bit-rate ranges)
There will be 9 conditions common to both test 1 and test 2. The 9 conditions will therefore be tested six times (3 labs each in test 1 and test 2). Thus they will have 24*6=144 subjects per condition.

For each test there will be 36 conditions common to all three labs, and there will be 72 subjects for each of these 36 conditions.

The remaining 108 conditions (36 cond*3 labs) will be tested at one lab only, and will thus have 24 subjects for each of these 108 conditions.

Labs: Assume 6 labs,
2 525/50 (NTSC) labs, labN1 and labN2
2 625/50 (PAL) labs labP1 and labP2
2 525/60 & 626/50 labs labNP1 and labNP2

Assuming that each lab can do both test 1 and test 2 (in a given video format), the labs could be assigned tests as follows (each test consisting of 72 cond/subj per lab):

LabN1	LabN2	LabP1	LabP2	LabNP1	LabNP2
Test1N	Test1N			Test1N	
Test2N	Test2N			Test2N	
		Test1P	Test1P		Test1P
		Test2P	Test2P		Test2P

Both tests should take 1.5 hours per subject, for a total of 36 testing hours worst case (with 24 subjects, assuming 1 subject at a time).

This test design assumes that all 504 (24 scenes * 21 HRC's) conditions will be tested. If the labs doing the testing can not commit to this amount of testing, then a partial block design will be needed.

This table assumes each lab can perform both tests (within a given video format), but this is not necessary as long as there are enough labs to cover each of these tests.   For example, if a lab can only conduct one test, then another lab can perform the second test as its only test, or in addition to the tests it is already assigned.

Matrix 1 Scene[1] (row)-HRC (column) matrix

HRCs sorted by quality (bit rate) -> increasing quality																				
					625/50 Video Format															
					525/60 Video Format															

[1] Scenes are arranged by video format. Within video format, scenes are arranged by coding difficulty with coding difficulty decreasing down the matrix.

Matrix 2 – condition selection for test 1N (rows are scenes[2], columns are HRC's)

1	2	3	4	5	6	7	8	9	10	11	12
C	1	2	3	C	1	2	3	C	1	2	3
3	C	1	2	3	C	1	2	3	C	1	2
2	3	C	1	2	3	C	1	2	3	C	1
1	2	3	C	1	2	3	C	1	2	3	C
C	1	2	3	C	1	2	3	C	1	2	3
3	C	1	2	3	C	1	2	3	C	1	2
2	3	C	1	2	3	C	1	2	3	C	1
1	2	3	C	1	2	3	C	1	2	3	C
C	1	2	3	C	1	2	3	C	1	2	3
3	C	1	2	3	C	1	2	3	C	1	2
2	3	C	1	2	3	C	1	2	3	C	1
1	2	3	C	1	2	3	C	1	2	3	C

[2]Within video format, scenes are arranged by coding difficulty with coding difficulty decreasing down the matrix.

c: Test 1 common conditions (note that the common conditions within 3 TBD HRC's will be common to both bit-rate range tests (Test1 and Test2).

1: Lab 1 conditions

2: Lab 2 conditions

3: Lab 3 conditions

As the test is currently designed, there are no repeated conditions (scene - HRC combination) within a given test. These repeated conditions will be used to check each subject for consistency. There will be most likely be 3-5 repeat conditions within a given test (or test tape). Thus, they should not significantly affect the number of testing conditions per subject (or lab). These repeated conditions need to be defined before test tapes are made.

- MOS: Mean opinion score
  DMOS: Difference mean opinion score; Source - Processed

8 PARTICIPATING LABORATORIES

Several laboratories have expressed an interest in conducting subjective tests: CCETT France, CSELT Italy, CRC Canada, ATTC USA, FUB Italy and RAI Italy. Other laboratories willing to volunteer time and resources can contact Laura Contin or Philip Corriveau.

9 SCHEDULES

Schedules for the subjective tests have yet to be determined.

10 DEFINITIONS

TEST SEQUENCES - sequences which have been selected for use by the ILSC

SOURCE SEQUENCE - an unprocessed Rec. 601 test sequence

PROCESSED SEQUENCE - a source sequence encoded and decoded according to a certain HRC

HYPOTHETICAL REFERENCE CIRCUITS (HRC'S) - conditions set at different bit rates, resolution, and method of encoding.

DEMO TRIAL - trial to familiarize the subject with the test structure

WARM-UP TRIAL - practice trials which are not included in the analysis

TEST TRIAL - trial consisting of source and processed sequences, ratings of which are included in the analysis

RESET TRIAL - trial after a break in viewing which are not included in the analysis

TEST TAPES - tapes containing randomized test trials

EDIT DECISION LIST - time code specifications for placement of test trials for the production of test tapes

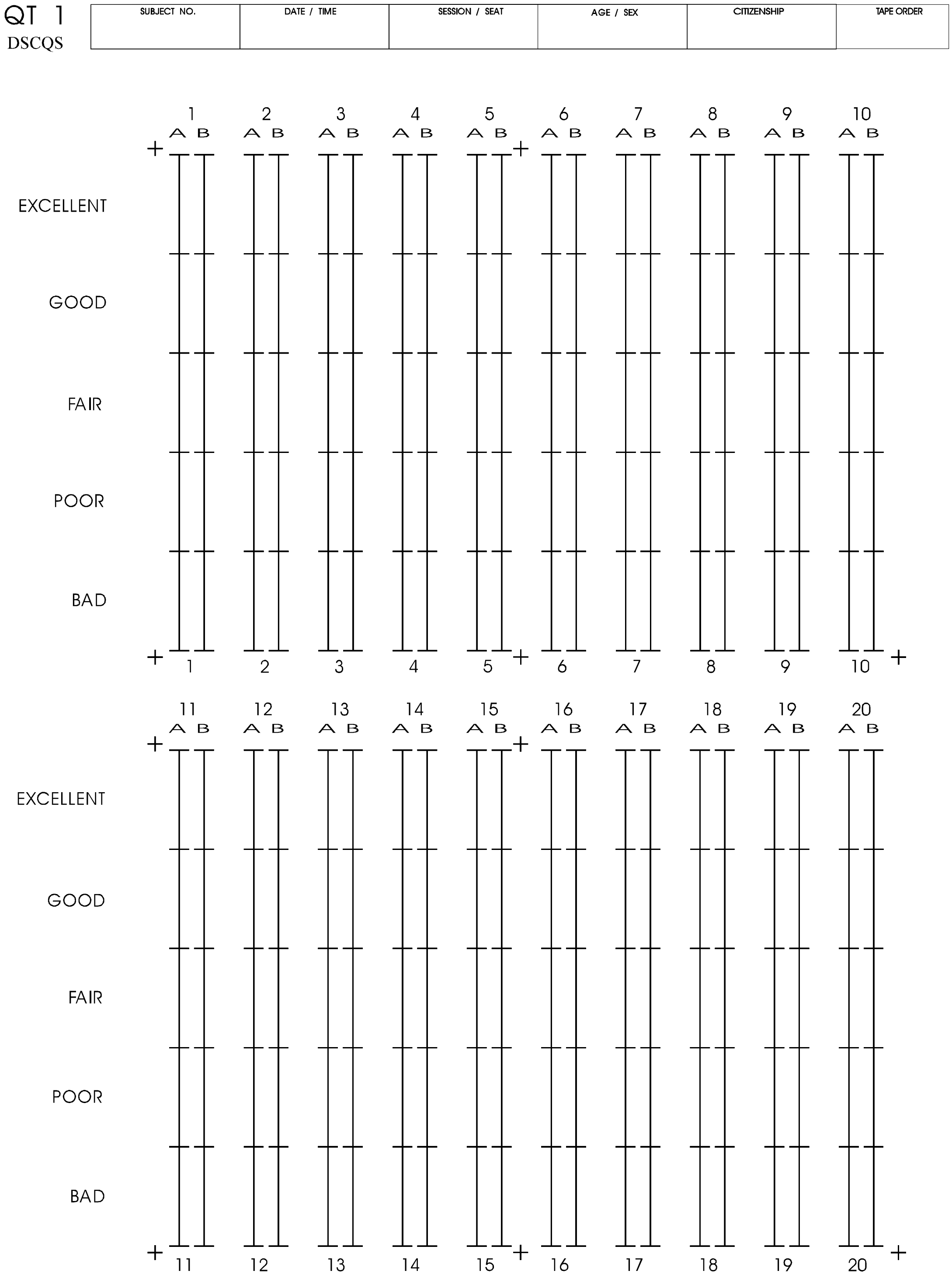
CONDITIONS - variables such as HRC's and sequence that are manipulated in this experiment

SESSION - a time period during which a series of test tapes is viewed by a set of subjects

CONTEXTUAL EFFECTS - fluctuations in the subjective rating of sequences resulting from the level of impairment in preceding sequences. For example, a sequence with medium impairment that follows a set of sequences with little or no impairment may be judged lower in quality than if it followed sequences with significant impairment.

ANNEX 1

Sample page of response booklet:



QT 1
DSCQS

SUBJECT NO.	DATE / TIME	SESSION / SEAT	AGE / SEX	CITIZENSHIP	TAPE ORDER