CONTRIBUTION TO T1 STANDARDS PROJECT

****

STANDARDS PROJECT: Analog Interface Performance Specifications for Digital Video Teleconferencing/Video Telephony Service

****

TITLE: Subjective Test Plan (Tenth and Final Draft)

****

EDITOR: A. C. Morton, AT&T Communications

SOURCE: Detailed Test Plan Ad Hoc Group and Data Analysis Ad Hoc Group

CONTACT:

A. C. Morton
AT&T Bell Laboratories
Room HO 3M-533 P.O.Box 3030
101 Crawfords Corner Road, Holmdel, NJ 07733-3030
908 - 949 - 2499

****

DATE: October 3, 1993

****

DISTRIBUTION: T1A1.5

****

ABSTRACT:

This is the subjective test plan required by T1A1.5 to complete its video performance specification project. The Detailed Test Plan Ad Hoc Group prepared this plan for the February 8, 1993 meeting, where version T1A1.5/93-014 R1 was originally accepted by the full Working Group.
In the process of developing the Data Analysis Plan, some of the calculations originally envisioned were deemed unnecessary. The Data Analysis Ad Hoc Group was then free to develop a more balanced sampling plan and make other improvements to the subjective test plan.
Revisions and new material from subsequent meetings of the Data Analysis Ad Hoc Group were adopted as version T1A1.5/93-014 R5 (on a technical basis) at the August 9, 1993 Working Group meeting. Some changes and additional appendices were adopted at the November 1993, January 1994, and July 1994 meetings of the Working Group.

Video Performance Standard Subjective Test Plan

1. INTRODUCTION

Working Group T1A1.5 is in the process of preparing an ANSI standard on video teleconference system performance measurement. The process includes steps to identify objective measures of video performance, to compare the objective measures with user opinion of video quality, and to select from the candidate measures those that are well-correlated with user opinion, as only these measures offer the desired information.

During the T1A1.5 meetings in October, 1992, agreement was reached on a set of 25 video test scenes and on 25 Hypothetical Reference Circuits (HRC). The test scenes, along with candidate objective test waveforms, have been assembled on D2 format video tape and played through the 25 HRC's. The 25 by 25 matrix results in 625 different test combinations, where a test combination is defined as the record of a single scene transmitted through a single HRC.

To correlate the candidate measures with a representative user's view of video quality, an estimate of the perceived quality of each test combination must be available. Working Group T1A1.5 formed an Ad Hoc Group to develop a detailed procedure for both subjective and objective testing. This document describes the subjective test procedure to be followed at each of the three test laboratories (Delta Information Systems, GTE Labs, and NTIA-ITS), in which the above test combinations become the stimuli for a video grading task.

The procedure will follow CCIR Recommendation 500-5 in general. This document identifies the specific sections and procedures to use, since there are many options within the Recommendation. Further, additional details specific to this test program will be defined here, such as the number of subjects required to view each test combination, and the division of test combinations between the three labs. In this way, the controlled conditions and test delivery will be determined and maximal consistency among the laboratories should result.

The GOAL of this Test Procedure is to estimate the distribution of opinion of video teleconference system users when presented with representative video sequences.

The following points represent the group's philosophy concerning the standard's development:

1. The intent of this process is to study the relative performance of a set of proposed objective video measures as predictors of subjective judgement.
2. The process is an evolutionary one, and the membership does not necessarily expect to reach the final set of measures in a single step, or single cycle through the process.
3. While negative contingencies or failures of the process may not be defined to the last detail, the membership believes that they have experience necessary to recognize when the results indicate failure. Further, that they are willing to

proceed with this research process with an understanding of the risks.

4. There is a need, raised here and previously by other members, to develop a document which describes the data analysis process step in some detail. This document should exist prior to the completion of the testing steps. (The completed data analysis plan appears in Section 5.)

5. It is recognized that the precision of the conclusions drawn as to the relationship between objective and subjective performance will be based on the precision of the basic subjective testing. The Working Group in this plan will assess the level of precision needed for the results to be compelling and to receive industry consensus.

In order to be useful in the standard development process, this procedure is also consistent with the Scope and Purpose of the Draft VTC Performance Standard (T1A1.5/94-107).

2. MAJOR TEST DESIGN ATTRIBUTES

This section contains the consensus position of the Ad Hoc Group on several major areas that required determination before any subjective tests could begin.

The general starting point was this list of design requirements:

1. Test a broad range of Hypothetical Reference Circuit types.
2. Use a broad range of Test Scenes.
3. Recruit an adequate number of viewers representing a well-defined target population.
4. Test as many of these combinations as feasible.
5. Adopt a partially balanced design which ensures that the quantities of interest are not confounded with unmeasurable sources of variation.
6. Use equal-probability sampling if possible.
7. Include appropriate quality checks.
8. Use CCIR Recommendation 500-5 as a guideline.
9. Use Digital Play/Record and editing to minimize generation loss.
10. Test the subjects for suitability (i.e. vision acuity).

2.1 Test Matrix

As stated earlier, agreement was reached on a set of 25 video test scenes and on 25 Hypothetical Reference Circuits (HRC). The 25 by 25 matrix results in 625 different test combinations, as shown in Table 1.

Table 1. TEST MATRIX

SCENES	HYPOTHETICAL REF. CIRCUITS							
	1	2	3	4	.	.	.	25
a	a1	a2	a3	...				a25
b	b1	b2	...					
c	c1	...						
d	...							
.								
.								
.								
y	y1	...						y25

2.2 Number of Test Subjects

The objective of these tests is to obtain an experimental Mean Opinion Score (MOS) for each test combination where the value obtained differs by no more than ± 0.2 opinion score points from the true mean value with 95% confidence. 30 test subjects are estimated as required. The sample size was determined using the following method.

First, the standard deviation, s, was estimated (for this experiment) by review of previous experimental results. Since the range of s was found to be large ($0.1 < s < 1.0$ on a five point grading task), a representative value for s was chosen. It is

$$s = 0.5$$

Using the following equation, we determined the necessary sample size to meet the 0.2 score point confidence interval requirement, e.

$$\pm e = \frac{qt(0.975, n) \times s}{\sqrt{n}}$$

where $qt(0.975, n)$ is a percentile of the Student's t distribution for double-sided confidence intervals at 95% and n is the unknown sample size.

When $n = 30$, $e = 0.186$ and the requirement is satisfied. For the test combinations where the sample standard deviation is more than 0.5, slightly larger confidence intervals will result.

Test labs must provide 30 viewer opinions for each test combination as a minimum after screening (see Section 2.6). Viewers are expected to rate all test combinations that are shown to them. The entire data set for the 30 viewers must be provided.

2.3 Target Population and Viewer Qualifications

Viewers selected for this experiment must have normal or corrected-to-normal visual acuity and color vision. These faculties will be tested prior to participation. Viewers that do not meet these requirements must not be included in the sample.

In keeping with the goal of the plan, the viewers should represent typical video telephony/teleconference (VT/VTC) system users. Ideally, they should be persons who

use these systems now, or envision using them in the next few years. The viewers should not be persons working directly on, or in support of, the design, sales, maintenance, or performance assessment of VT/VTC systems or services.

There is one additional qualification. 50% of the viewers at each lab should have some experience with video conferencing. Some allowances may be made for recruiting difficulties.

2.4 Stimuli Presentation and Voting Method

CCIR Rec. 500-5 describes several test methodologies for subjective assessment of television pictures. The grading scale used determines the user's measure of scene rendition. The ANSI standard for expressing video performance will make reference to this scale.

The Working Group agreed on the double-stimulus/impairment scale method as described in section 2 of CCIR 500-5, with some modifications. The modifications are:

1. The rating scale used by subjects will not show the numerical values 5 through 1. These values (5 = Imperceptible) will be assigned during data entry.
2. The reduction of time intervals as defined in the presentation of test material (Figure 2 of Rec. 500-5), to allow 9 seconds to view the reference scene, a 3 second gap, 9 seconds to view the impaired scene, and a 9 second voting interval.
3. A mid-gray level of 50 IRE will be used in the interval between pairs of scenes and during the voting interval.

2.5 Voting Forms

Appendix B contains a sample voting form which is the result of collaborative effort on the part of many working group members.

2.6 Quality Checks

There are four necessary quality checks:

1. Viewer reliability will be tested through repetition of one test combination in every session. Viewers will be disqualified if their grades differ by more than 2 opinion score points (of the 1 through 5 scale) for the repeated test combination in any test session. Combinations selected for repetition will come from HRC's with either 384 kbps or 768 kbps transmission speed, so as to avoid combinations whose expected average rating scores are near either end of the rating scale.
2. Viewer reliability will also be tested through the distribution of Null HRC conditions among all test labs. Each test session will contain at least 1 Null combination. Viewers will be disqualified if they grade the Null combination at 3 or less (of the 1 through 5 scale) in any test session.
3. As many as 2 missing ratings will be tolerated per viewer. If any missing rating is on a quality check combination, then the viewer will be disqualified.

4. Lab-to-lab consistency will also be tested through repetition of 75 combinations at each of the labs. See the following section.

2.7. Partially Balanced Sampling Plan

2.7.1 Allocation of HRCs and Scenes to Video Tapes

Setting aside four HRC's for special treatment, 21 HRC's were combined into a partially balanced design described here.

Considerations of viewer burden allows us to show only about one-third of the possible test combinations to any single viewer, spread across several sessions. It was therefore decided to create three sets of viewing tapes (designated as the "Red", "Green" and "Orange" sets, or R, G and O, for short), each set of tapes containing all the scenes, but only one-third of the HRC's, in all possible combinations. Any given viewer will see exactly one set of tapes.

Referring to the HRC's by number in accordance with document T1A1.5/92-174 (see Appendix D), we allocated the HRC's to the sets of viewing tapes as follows:

Red Tape Set:	1, 4, 7, 8, 13, 15, 19, 20, 22, 24
Green Tape Set:	2, 5, 6, 10, 14, 15, 16, 17, 20, 23
Orange Tape Set:	3, 4, 9, 11, 12, 17, 18, 20, 21, 25

This allocation was guided by a desire to include the full range of video performance in each tape set, observing that the HRC's fall into 9 general types according to their engineering descriptions (see Table 3). In particular:

— An even division was achieved among codec types and transmission rates.

— Each tape set has 2 or 3 proprietary HRC's, 1 or 2 QCIF HRC's, 5 CIF HRC's (not counting the Null or VHS HRC).

— Each tape set contains one HRC with transmission errors.

Additionally, the four remaining HRC's that had been set aside at the start were given special treatment by being included in more than one tape set, each. The purpose is to allow post-hoc calibration checks between the tape sets. They were allocated to the tape sets as follows:

TAPES	HRC No.	Description
R,G,O	20	Identical Px64 Codecs at 384 kB/s
R,G	15	Identical Px64 Codecs at 112 kB/s
G,O	17	Different Px64 Codecs at 128 kB/s
R,O	4	Vector Quantiz. Codec at 128 kB/s

2.7.2 Allocation of Tape Sets to Testing Labs

Three different video labs volunteered to participate in the subjective viewing and data gathering phase of the study. Early thinking called for sending each of the three color-coded sets of tapes to just one lab, which would have allocated each HRC to just a single lab -- except for the four HRC's that are repeated across 2 or 3 labs. This plan had a

certain appeal in terms of its logistic simplicity, but it has the drawback that any inter-laboratory differences would have to be assessed, and possibly corrected for, using only a tiny fraction of the test material and data.

Other similar studies suggest that inter-laboratory differences might indeed occur -- either because of uncontrollable differences in the physical conditions of the test set-up, or because of differences in the sampled populations of viewers at the three locations. We did not want to be left with a set of data in which quantities of primary interest (HRC quality ratings) are essentially confounded with unmeasurable, irrelevant, and possibly inexplicable factors (collectively called "inter-laboratory differences"), merely because of weaknesses in the sampling design.

So the initial plan was abandoned in favor of a more balanced design, in which all of the color-coded tape sets are sent to each viewing lab. Every lab is instructed to divide its 30 test subjects randomly into three teams of 10 subjects each, for viewing the three sets of tapes. Thus, each lab will assemble a "Red Team", a "Green Team" and an "Orange Team". In the overall data set, the "Red Cohort" will be the union of the Red Teams from the three labs, and will be spread in equal numbers across the labs, and so on for the other colors. Schematically, we have:

Lab	Tapes	Viewers	Alternates
X	Red	X1 ... X10	X11 ... X20
	Green	X21 ... X30	X31 ... X40
	Orange	X41 ... X50	X51 ... X60
Y	Red	Y1 ... Y10	Y11 ... Y20
	Green	Y21 ... Y30	Y31 ... Y40
	Orange	Y41 ... Y50	Y51 ... Y60
Z	Red	Z1 ... Z10	Z11 ... Z20
	Green	Z21 ... Z30	Z31 ... Z40
	Orange	Z41 ... Z50	Z51 ... Z60

(N.B. the viewer sequence numbers above are their numbers *after* being randomized according to instructions in Section 2.9. The numbers for alternate viewers may not be completely used.)

Since the quantities of primary interest in the data analysis will be summaries across the cohorts, any laboratory-specific factors affecting judgements will be neutralized by being equally spread out across all results.

This plan turns a potential liability (inter-laboratory differences) into a strength: the pooled data set from the three labs can be regarded as a properly stratified random sample from a target population that is an equal-probability mixture of the target populations realized at the three labs. In this way, by going to three labs instead of one for test subjects, we are likely to broaden the scope of our sampled population and make it more

representative of the true potential market for video teleconferencing in the country.

This plan has the further desirable property that any given HRC has the same probability of being viewed by every test subject in the study, which justifies the use of unweighted averages across cohorts as efficient and unbiased estimates of population parameters.

Moreover, this design will also allow us to study the inter-laboratory differences themselves -- with a view to clarifying our understanding of the target sub-population that were *actually* sampled.

2.8 Stimuli Presentation Order

2.8.1 Number of Test Sessions

It was observed that with some slight modifications to the test method outlined in CCIR Rec. 500-5, the required number of stimuli that we must present to each viewer can be accommodated in 4 viewing sessions of 32.5 minutes each -- exceeding the CCIR recommendation by just 2.5 minutes. Each session will consist of

10 HRC's * 25 Scenes * 0.5 min/seq. = 125 min of testing

125 min / 4 sess = 31.25 min/sess

To each session, we add 1 minute for two additional calibration checks (one repeated combination and one Null combination). This gives an average of 32.25 min per session, or two sessions with 32.5 minutes and two with 32.0 minutes.

2.8.2 Randomization Within Test Sessions

Principles of good experimental design require that all the test combinations shown to a viewer be randomly permuted over the viewer's four session tapes. This permits each session to exhibit a full range of video quality and mitigates artifacts due to presentation order (learning or fatigue effects, adjacency effects, etc).

Ideally, we might want each viewer to be presented with an independent randomization of the stimuli, but we are constrained by the need to prepare each session as a pre-edited 1/2 hour video tape, so we plan to use the same 1/2 hour tapes (and hence the same randomization sequence for each 1/2 hour tape) for all the viewers presented with the same selection of HRC's. The presentation order for the four session tapes will be randomized as per section 2.9.1.

The order of presentation of the stimuli can influence the opinion of the evaluators in subtle ways. Therefore, although a randomized order of presentation is necessary, it may not be sufficient. Following the CCIR guidelines, we arranged for consecutive stimuli to be dissimilar on each of the two design dimensions, meaning that not only the pictorial content but also the transmission impairments caused by the characteristics of the HRC's varies. This was achieved by making some minor modifications to the randomization process which take into account the grouping of HRC's into 9 types, and a grouping of the 25 scenes into 5 categories with similar pictorial content, shown in Table 2.

Table 2 identifies each scene by its short process name and the lower case letter used in the test matrix on Table 1.

Table 2. SCENE CONTENT CATEGORIES

CONTENT CATEGORY	DESCRIPTION	SCENE NAMES & LETTERS
A	One person, mainly head and shoulders	vtclnw(f), susie(j), disguy(k), disgal(l)
B	One person with graphics and/or more detail	vtcmp(a), vtc2zm(b), boblec(e), smity1(m), smity2(n), vowels(w)
C	More than one person	3inrow(d), 5row1(g), intros(o), 3twos(p), 2wbord(q), split6(r)
D	Graphics with pointing	washdc(c), cirkit(s), roadmap(t), filter(u), ysmite(v), inspec(x)
E	High object and/or camera motion (Examples of Broadcast TV)	flogar(h), ftball(i), fredas(y)

Table 3 divides the 25 HRC's into 9 groups according to the transmission quality and type of impairments that are to be expected. Some revisions may be desirable after the processed tapes have been reviewed.

Table 3. REFERENCE CIRCUIT GROUPS

GROUP NO.	CIRCUIT DESCRIPTION	HRC NO.
1	High Quality	1-3
2	Vector Quantization, medium rate	4-5
3	Proprietary, low to medium rate	6-7
4	Proprietary, medium to high rate	8-10
5	QCIF, low rate	11-13
6	QCIF, medium rate	14
7	CIF, low rate	15-18
8	CIF, medium rate	19-21
9	CIF, high rate	22-25

Each processed test scene, also called a test combination, was then assigned a number and letter code (such as 5-B) roughly categorizing pictorial content and transmission circuit characteristics.

In terms of these categories and this notation, the randomization process was performed in the following steps, using sampling without replacement, subject to certain constraints:

1. All test scenes processed through the 10 HRC's assigned to one set of viewing tapes were put into a pool (the number 10 includes the 7 exclusively assigned to that set, plus the 3 that are shared across sets).
2. Randomly, a test scene was pulled and its code checked.
3. If both number and letter were different from the preceding scene, it was accepted.
4. If either the number or letter were the same, it was returned to the pool and another scene was pulled, until one was found that was accepted.
5. This was continued until all spaces on the tape (either 64 or 65) were filled.
6. The whole process above was then continued for the next session tape in the set, using the remaining combinations.

This selection process results in four well-randomized tapes for each HRC set. We anticipated some difficulties in satisfying the adjacency constraints toward the end when only a few scenes remained in the pool. Judicious exchange with previously assigned scenes made it easy to solve this problem.

With 64 or 65 stimuli per tape and only 10 HRC's per tape, the above described randomizing process produced a healthy balance of HRC's across the four session tapes in each set.

Special attention was given to the scenes which are to be used for quality checks since their number is held to a minimum to avoid excessive disqualifications. The Null circuit scenes will be judged by viewers primarily in terms of resolution and color fidelity, since motion rendition is a minor factor. Therefore, scenes washdc(c), flogar(h), cirkit(s), and rodmap(t) were prime candidates for this purpose. One was assigned to each session tape. Scenes selected for repetition were typical and average ones, especially those with the content/quality code 8-C or 9-D.

The location of Null circuit and repeated scenes on the tape were not determined by the randomization process described above. Instead, locations were picked judiciously, by hand, to ensure that each session tape contains one of each and that all other constraints were satisfied. These locations were different on each session tape.

The above randomization prescriptions produced a satisfactory sequence of scenes on each tape, but one further level of randomization was adopted to further neutralize potential order-of-presentation artifacts -- a block randomization achieved by presenting the four session tapes in a different randomized order to different sets of 1, 2, or 3 viewers, as described in Section 2.9.1.

2.9 Procedures for Randomized Tape Viewing and Selection of Viewer Groups

The following guidelines for the subjective testing laboratories specify how to divide the pool of viewers into session sub-teams and what tape presentation order to use for each of the sub-teams. The intent of these guidelines is to minimize systematic differences which could lead to biases in HRC ratings. Whenever possible, the testing laboratories should use these guidelines. Any exceptions to these guidelines will be recorded (Appendix G gives an alternative method for random selection of viewer groups that will

be used at one or more labs). Here the labs are identified as X, Y, Z and the tapes as R1, R2, R3, R4; G1, G2, G3, G4; O1, O2, O3, O4.

2.9.1 Random Ordering of Session Tapes

To allow for the possibility of tape sharing between the labs, it is assumed that each set of four tapes will be viewed in succession without interruption by other tapes. The order which teams view the tapes is determined randomly subject to balance. Using Tables of Random Permutations by L.E. Moses and R.V. Oakford (Stanford University Press, 1963) yields the following chronological order of the tape sets:

Table 4. CHRONOLOGICAL ORDER OF TAPE SETS

Lab X:	G O R
Lab Y:	O R G
Lab Z:	R G O

Costs permitting, a complete set of tapes will be prepared for each lab to facilitate scheduling. Each of the 9 teams of viewers given above must have at least 10 viewers. The four tapes shown to each of the nine teams are to be ordered in a random fashion. Since each session can have at most 3 viewers, at least 4 random tape orderings for each group of 10 viewers is possible. For all 9 groups, this gives a total of 4 x 9 = 36 orders. There are 4!=24 permutations of the integers 1,2,3,4. It seems reasonable to use all 24 permutations and include random duplicates of 12 of them. This is done by using Table 1 of Moses and Oakford, which consists of 960 permutations of 1,2,...,9, ignoring the numbers 5 through 9, copying the successive orders of 1,2,3,4, deleting any order that arises more than twice, and accepting exactly 12 duplicates as they arise until 36 are obtained:

Table 5. TAPE ORDER PERMUTATIONS

1423	2134	3214	1432	3421	4321	[1423]	4231	1342
1234	2143	3412	3124	4123	[1234]	2413	2314	[2134]
3142	[2314]	[1342]	4213	[3124]	[3412]	[3142]	[2413]	[2431]
[4231]	[4321]	1243	2431	2341	1324	4132	4312	3241

Hence the orders in which the tapes are to be viewed are as follows, proceeding from left to right and line by line:

Table 6. VIEWING ORDER FOR SUB-TEAMS BY LAB

Lab X:

G1	G4	G2	G3	G2	G1	G3	G4	G3	G2	G1	G4	G1	G4	G3	G2
O3	O4	O2	O1	O4	O3	O2	O1	O1	O4	O2	O3	O4	O2	O3	O1
R1	R3	R4	R2	R1	R2	R3	R4	R2	R1	R4	R3	R3	R4	R1	R2

Lab Y:

O3	O1	O2	O4	O4	O1	O2	O3	O1	O2	O3	O4	O2	O4	O1	O3
R2	R3	R1	R4	R2	R1	R3	R4	R3	R1	R4	R2	R2	R3	R1	R4
G1	G3	G4	G2	G4	G2	G1	G3	G3	G1	G2	G4	G3	G4	G1	G2

Lab Z:

R3	R1	R4	R2	R2	R4	R1	R3	R2	R4	R3	R1	R4	R2	R3	R1
G4	G3	G2	G1	G1	G2	G4	G3	G2	G4	G3	G1	G2	G3	G4	G1
O1	O3	O2	O4	O4	O1	O3	O2	O4	O3	O1	O2	O3	O2	O4	O1

2.9.2 Random Selection of Viewer Groups

Each lab will assemble a pool of viewers, about one-third of whom will view the R tapes, another third the G tapes, and the remaining third the O tapes. Since not more than three viewers will participate at any session, there will be at least four sub-teams of viewers of R tapes, four of G tapes, and four of O tapes at each lab, 12 non-overlapping sub-teams at each lab. The following discussion assumes sub-teams of size three. Even if all members of a sub-team do not view the tapes at the same time, they must view them in the same order.

The viewers shall be assigned to sub-teams at random to avoid systematic differences between the R, G, and O teams, which could lead to biases in HRC ratings. Each lab shall list its viewers in alphabetical order and assign the numbers 1,2,... in that order.

Three tables of random permutations of 1,2,...,50 from Moses and Oakford are shown below. Lab X is to use the first permutation, Y the second, and Z the third, going down the columns and disregarding any numbers beyond the available number of viewers. If Lab X has 36 viewers, then its first sub-team has viewers No. 8, 3, 5; its second No. 13, 10, 27; its third No. 34, 12, 25; its fourth No. 14, 4, 23; its fifth No. 15, 18, 2; etc.

This means that the first sub-team at Lab X, viewers 8, 3, 5, will view the G tapes in the order G1 G4 G2 G3; the second sub-team of Lab X, viewers 13, 10, 27, will view the G tapes in the order G2 G1 G3 G4; the third sub-team in the order G3 G2 G1 G4; and the fourth sub-team in the order G1 G4 G3 G2. The fifth sub-team at Lab X will view the O tapes in the order O3 O4 O2 O1, etc., up to the twelfth sub-team at Lab X, which will view the R tapes in the order R3 R4 R1 R2.

Similarly Lab Y will choose 12 sub-teams according to the second permutation. If it has 37 viewers available, its first sub-team of three consists of viewers No. 24, 16, 36; its second, No. 21, 15, 2; its third, No. 11, 20, 31; etc.

Similarly Lab Z will choose 12 sub-teams according to the third permutation. If it has 36 viewers available, its first sub-team consists of viewers No. 22, 9, 15; etc.

Table 7. RANDOM VIEWER ASSIGNMENT TO SUB-TEAMS

Lab X

8	27	4	2	26	48	38	46	9	39
3	34	23	21	6	50	37	40	16	19
5	12	15	7	32	49	30	42	11	31
13	25	18	43	28	22	29	41	1	35
10	14	44	20	36	47	17	45	33	24

Lab Y

24	2	31	19	46	17	37	13	28	33
16	41	3	43	1	7	18	30	39	27
36	11	10	8	44	26	12	23	47	34
21	49	50	38	25	4	5	32	9	40
15	20	6	45	22	29	14	42	35	48

Lab Z

22	15	31	12	5	24	14	29	45	30
39	44	43	41	16	8	21	35	49	37
40	18	28	47	11	33	36	20	42	26
9	38	7	17	46	25	23	2	4	19
48	13	50	34	6	10	32	27	3	1

3. PRE-TEST PROCEDURES

1. All viewers will complete a pre-test questionnaire (See Appendix C).
   Note: The subject demographic data will be examined only if the need arises, such as if differences emerge in the Lab to Lab calibration checks and possible reasons are sought for the differences. However, the 3 test labs will need to examine the question on video telephony usage, to ensure that they have met the sample experience requirement.
2. All viewers will complete vision acuity tests at the (6H) viewing distance, and color tests at the distance recommended in the color test procedure. The specific test for acuity is a Graham-Field Catalog # 13-1240 single letter identification chart, or equivalent, and modified for use at 6H distance. The specific test for color vision is the Pseudo-Isochromatic Plates for Testing Color Perception, as supplied by Beck Engraving Co., for example.
3. At each session, Labs will record the session number, tape number, viewing position(s), and time of day.
4. Instructions must be standardized and delivered on tape. Appendix A gives the text of the Instructions.
5. There will be a practice session, consisting of 6 cycles through the viewing and voting process. The range of quality displayed to the subjects will represent nearly the full range of quality in the experiment and will also contain one combination with transmission errors. In the spirit of conforming with CCIR 500-5, there will be a 2 minute break between practice and test sessions.
6. During subsequent sessions, the subjects will again view the practice tape sequence prior to the new material. There should be no need to practice voting in these sessions.

4. TEST SESSIONS

4.1 Considerations

1. Test scene numbers must be announced prior to the sequence.
2. Audio cues to vote will be included.
3. Use voting forms as shown in Appendix B.

4.2 Session Scheduling

The scheduling of the test sessions (time of day, rest between sessions, etc.) can affect the test results. Compliance with CCIR 500-5 and availability of test personnel can produce significant constraints on the time at which each test is performed.

Scheduling of individual test sessions, in terms of the number of sessions per week per subject is left to each individual lab's discretion. There will be a 15 minute break between sessions in a pair, and a 1 minute break during each session. There will be at least 1/2 day rest between pairs of test sessions.

Test personnel should not change between sessions.

4.3 Viewing Conditions

In preparing the viewing conditions, the test labs will use Section 2 of CCIR 500-5 as a guideline. All viewing will be conducted at a distance equal to 6 picture heights (6H).

The viewing monitor used at each test lab will be the SONY BVM-1910 or equivalent.

The format of the session tapes will be Betacam SP.

The SMPTE color bars alignment signal will be available on tape.

5. INITIAL ANALYSIS OF SUBJECTIVE DATA

The purpose of this section is to define the data processing and initial analysis that will be performed on the data obtained from the subjective tests. The steps described in subsections 5.1 through 5.6 are to be performed at each lab on the data obtained at that lab.

5.1 Data Entry

The inputs to the Data Analysis function are the score sheets filled in by the test observers (viewers) at each of the 3 labs. At each lab, there will be 30 primary viewers, allocated into 3 teams of 10 each, and possibly some number of alternates or replacements for each team. The teams are coded as Red, Green and Orange, corresponding to the color-coding of the video tapes they will view. The tapes come in sets of 4 for the 4 sessions that each viewer will attend. For each viewer and each session there will be 64 or 65 opinion scores. Each score will be assigned an integer from 1 to 5. For the total experiment there will be 23,220 opinion scores (not counting disqualified viewer and surplus alternates). The original score sheets will be duplicated, and the original sheets will be stored at a central location (place TBD). All further processing will be performed on the copies.

The data from each score sheet will be keyed into a computer. The data on each sheet consists of the test observer ID, the lab ID, the session number, and 64 or 65 opinion scores. Data will be keyed in by two independent operators, and a computer program will verify that the entered data is consistent.

Laboratories should coordinate the keying procedure to simplify interlab exchange of data (e.g., ASCII).

5.2 Convert Presentation Sequence

The randomization sequences for each lab, as shown in Appendix F, will be made available by Delta Information Systems in WordPerfect format. These sequences will be converted by each lab to a format that can be used for its computer program.

5.3 Perform Null and Repeat Quality Checks

Each lab will analyze its own data to determine the quality of the scores given by each test observer.

In each test session there is a Null (designated #0 in Appendix F) inserted to test the observer. If on any of these Nulls the observer gives a score of 3.0 or less, the observer will be disqualified.

In each test session one combination of HRC and scene is repeated. If for any repeated combination the absolute difference between the two opinion scores given by a test observer is 3.0 or more, the observer will be disqualified.

Alternate viewers for each team (if any) are initially marked as disqualified. Some number of them will subsequently be marked as qualified if they pass the two quality checks and if they are needed to replace disqualified primary team members.

5.4 Reformat the Data

Using the keyed-in score data and the presentation sequences for each session, each lab will use a computer spreadsheet program to reformat the data into a series of tables as shown below.

In each session, there is one combination that is repeated twice; the score from the first occurrence shall be entered into the main body of the table, and the second shall be entered into the "REPEAT" column.

5.5 Compute Summary Statistics Per HRC

For each of the reformatted tables (HRCs) obtained from Section 5.4, the following calculations are performed by computer at each lab:

The average and sample standard deviation of each column (scene) are calculated across all qualifying rows, and displayed to two decimal places in the table.

The standard error of the mean of each column is calculated as the standard deviation calculated above divided by the square root of the number of qualifying values in the column, and displayed.

LAB _____ HRC NO. _____

TEAM	-- VIEWER -- ID	QUALIFY?	SCENE a	b	c	d	e	...	w	x	y	-- REPEAT -- SCENE	SCORE
RED	1	1											
	2	1											
	...												
	10	1											
(ALT.)	11	0											
	12	0											
	...												
	20	0											
GREEN	21	1											
	22	1											
	...												
	30	1											
(ALT.)	31	0											
	32	0											
	...												
	40	0											
ORANGE	41	1											
	42	1											
	...												
	50	1											
(ALT.)	51	0											
	52	0											
	...												
	60	0											
MEAN													
S.D.													
S.E.													
MAX													
MIN													

The maximum and minimum qualifying value in each column are determined and displayed.

5.6 Lab-to-Lab Analysis

Up to this point, all analysis will be performed by each lab on the data that it has gathered. From this point on, the data will be analyzed as a whole. The results obtained from Sections 5.4 and 5.5 will be distributed to all concerned parties for further analysis.

Qualifying data from the various labs will be studied for interlab consistency, and pooled in a statistically suitable fashion, to produce the tabulation below for use in the next (objective) phase of the study.

Plans for further data analysis will be presented in future T1A1.5 contributions.

SUBJECTIVE TEST RESULTS

		SCENE a b c d e ..... x y
HRC		
1	mean s.d. s.e. max min	
2	mean s.d. s.e. max min	
3		
25	mean s.d. s.e. max min	

NOTE: s.d. estimate of standard deviation of observer population
s.e. estimate of standard error of mean
max largest observed opinion score
min smallest observed opinion score

6. AD HOC GROUP MEMBERSHIP

The complete list of Detailed Test Plan Ad Hoc Group members and Data Analysis Plan Ad Hoc Group members can be found in Appendix E.

APPENDIX A - Instructions to Test Subjects

Instructions for Initial Session

The ANSI Standards Committee on Telecommunications is conducting a study to determine how different video telephony systems affect the delivered quality. Today we are asking you to help us measure the quality of the video scenes you are about to see.

We are not asking you to rate the content of the scenes, the artistic composition nor the quality of the acting. Rather, we are asking for your rating of the quality of the video image itself.

You will be shown two versions of the same video scene. The first version will be the original scene. The second version will be the original passed through a video transmission system. You are asked to rate the difference in quality that you perceive between the two versions. As shown on your rating form, please score the differences as either Imperceptible (when you cannot see a difference between the first and second versions), Perceptible but not Annoying, Slightly Annoying, Annoying, or Very Annoying.

You should mark your rating form as indicated on the practice sheet. Note that response ovals for each scene pair are arranged in a vertical column below the scene number. Mark only one oval for each scene pair.

Remember, there are no right and wrong answers. We are interested in how you, personally, perceive the difference between the two versions. It is not necessary to think long about your answer. However, please watch the entire scene before scoring. Please do not discuss the scenes with your fellow viewers. Your first reaction is what we wish you to record.

We now begin the Practice Section. Use the your rating form labeled "Practice Section" and mark the appropriate oval for the following six pairs of scenes. The practice session displays representative examples of video quality in the test.

The announcer will give the number of each pair of scenes before it appears. After viewing each scene pair, the announcer will ask you to rate the video quality.

We now begin the rating process for the Practice Section.

Here is scene 1. (speak quickly)

.
.
.

Please score scene 1.

...
...
...
...

Here is scene 6.

.
.
.

Please score scene 6.

That completes the Practice Section. We will begin the First Section in a few moments.

Instructions for Follow-on Sessions

Welcome back to the Video Viewing Lab.

Today's session will be similar to your last session with us.

You will be shown two versions of the same video scene. The first version will be the original scene. The second version will be the original passed through some video system. You are asked to rate the difference in quality that you perceive between the two versions. As shown on your rating form, please score the differences as either Imperceptible (when you cannot see a difference between the first and second versions), Perceptible but not Annoying, Slightly Annoying, Annoying, or Very Annoying.

Please watch the entire scene before scoring, and do not discuss the scenes with your fellow viewers. Your first reaction is what we wish you to record.

The announcer will give the number of each pair of scenes before it appears. After viewing each scene pair, the announcer will ask you to rate the video quality.

We now begin the Practice for the First Section. You should mentally score the scenes in this sequence, but do not mark your rating form at this time.

Instructions for Viewing Session Tapes

We now begin the rating process for the First Section.

Here is scene 1.

.
.
.

Please score scene 1.

...
...
...
...

Here is scene 30.

.
.
.

Please score scene 30.

That completes the First Section. We will begin the Second Section in a few moments.

(60 second break)

We now begin the rating process for the Second Section.

("Here is scene 31" and so on to scene 64 or 65)

That completes the Second Section.

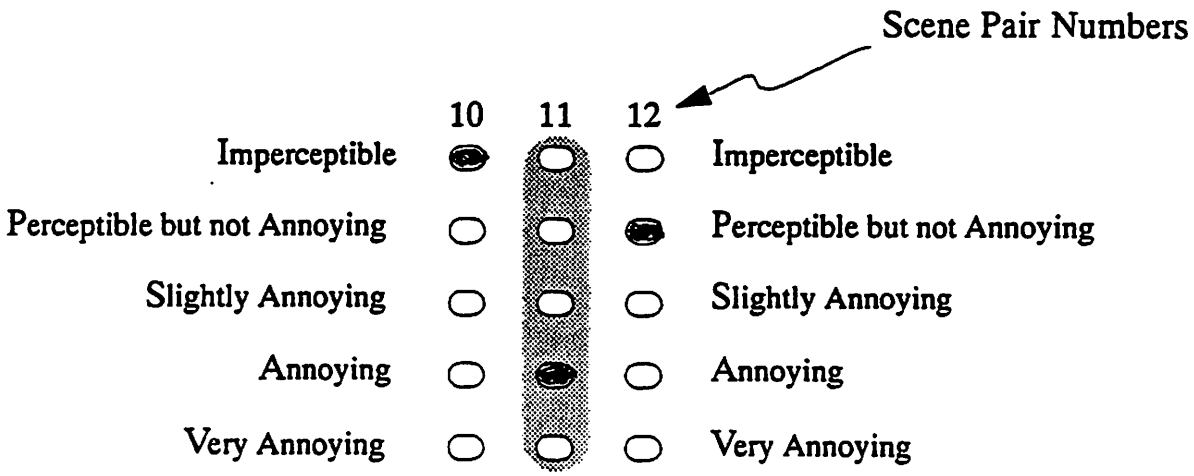
This concludes today's video quality rating session. We look forward to seeing you again for the next video quality rating session.

APPENDIX B - PRACTICE and VOTING FORMS

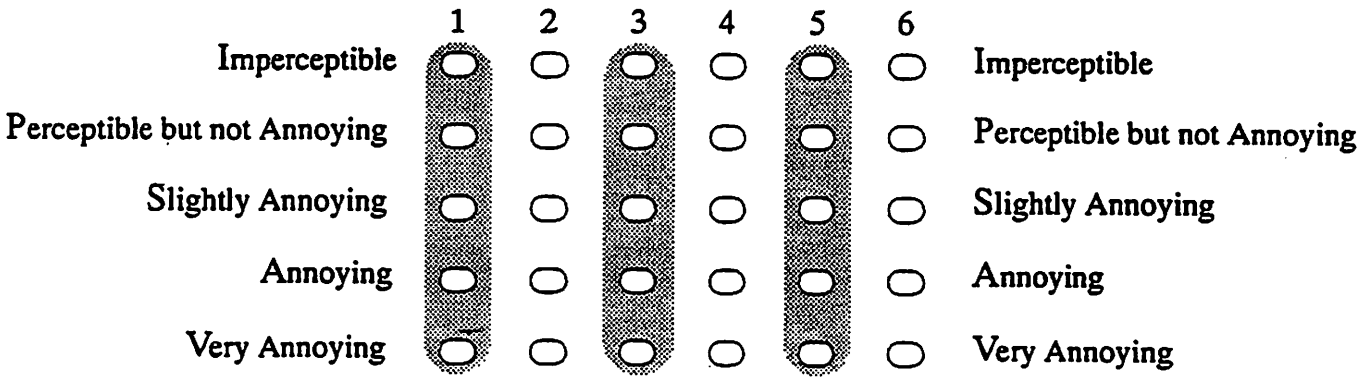
Viewer Rating Practice Form

You will be shown two versions of the same video scene. The first version will be the original scene. The second version will be the original passed through some video system. Rate the difference between the two versions by marking one of the five ovals. See the examples below which show typical markings for 3 scene pairs.

Marking Examples



Practice Scenes



APPENDIX B - PRACTICE and VOTING FORMS

	1	2	3	4	5	6	
Imperceptible	○	○	○	○	○	○	Imperceptible
Perceptible but not Annoying	○	○	○	○	○	○	Perceptible but not Annoying
Slightly Annoying	○	○	○	○	○	○	Slightly Annoying
Annoying	○	○	○	○	○	○	Annoying
Very Annoying	○	○	○	○	○	○	Very Annoying

Date _____ Time _____ Evaluator No. _____ Viewing Position _____ Tape No. _____

Signature _____

(Name)

APPENDIX C - PRE-TEST QUESTIONNAIRE

Please answer the following questions: NAME_____

1. Circle your age range:
   10-19 20-29 30-39 40-49 50-59 60-69 70-79 80-89

2. Circle your Gender:
   Male Female

3. Indicate your occupation according to industry and job: (Mark one in each column)

INDUSTRY	JOB
_____Accounting/Legal/Consulting	_____Administrative/White Collar
_____Advertising/Public Relations	_____Clerical/Support
_____Agriculture/Forestry	_____Driver
_____Broadcasting/Newspapers	_____Executive/Managerial
_____Construction/Contracting	_____Farmer/Forester
_____Education	_____Homemaker
_____Electronics/Computers	_____Industrial Worker
_____Engineering/Architecture	_____Laborer
_____Entertainment	_____Maintenance
_____Finance/Insurance	_____Owner/Operator
_____Government	_____Police/Military
_____Health Care/Social Services	_____Professional
_____Manufacturing/Printing	_____Rancher/Fisher
_____Military	_____Sales/Marketing
_____Mining/Oil/Gas	_____Secretarial
_____Personal/Business Services	_____Semi-Professional Services
_____Real Estate	_____Skilled Trades
_____Religion	_____Student
_____Restaurants/Lodging	_____Technical
_____Retail	_____Unemployed
_____Retired	
_____Security/Services/Police	
_____Telecommunication/Utilities	
_____Transportation	
_____Wholesale	
_____None	

4. Have you previously participated in video telephony, or in a video teleconference? (Mark one answer)

   — Yes, within the last 2 years.

   — Yes, but not within the last 2 years.

   — No.

APPENDIX D - TABLE OF HYPOTHETICAL REFERENCE CIRCUITS

This table first appeared in document T1A1.5/92-174. As permited in their charter, the Testing Ad Hoc Group (H. Meiseles, Vyvx, Chair; S. Gallaher, Vyvx; A. Morton, AT&T Communications) modified the table slightly to comply with the limitations of the available equipment. The modified version appears below.

HYPOTHETICAL REFERENCE CIRCUITS

HRC	Algorithm (vendor)	Resolution	Total, Kbps	Audio, Kbps	Video, Kbps	Coding Mode	Frame Rate	FEC	Burst Errors
1	Null	-	-	-	-	-	-	-	Off
2	VHS	-	-	-	-	-	-	-	Off
3	Proprietary	V.High	45,000	-	-	-	-	-	Off
4	Proprietary	Med.	128	-	-	VQ	-	-	Off
5	Proprietary	High	336	-	-	VQ	-	-	Off
6	Proprietary	Med.	112	-	-	-	-	-	Off
7	Proprietary	Med.	384	-	-	-	-	-	Off
8	Proprietary	Med.	768	-	-	-	-	-	Off
9	Proprietary	High	768	-	-	-	-	-	Off
10	Proprietary	High	1536	-	-	-	-	-	Off
11	H.261(diff)	QCIF	128	56	70.4	INTER+MC	-	On	Off
12	H.261(same)	QCIF	128	56	70.4	INTER	10*	On	Off
13	H.261(same)	QCIF	168	48	118.4	INTER+MC	-	On	Off
14	H.261(diff)	QCIF	384	56	326.4	INTER+MC	-	On	Off
15	H.261(same)	CIF	112	48	62.4	INTER+MC	-	On	Off
16	H.261(same)	CIF	128	56	70.4	INTER+MC	-	On	Off
17	H.261(diff)	CIF	128	48	78.4	INTER+MC	-	On	Off
18	H.261(same)	CIF	168	48	118.4	INTER+MC	-	On	Off
19	H.261(same)	CIF	256	56	190.4	INTER+MC	15*	On	On
20	H.261(same)	CIF	384	56	326.4	INTER+MC	-	On	Off
21	H.261(same)	CIF	384	56	326.4	INTER+MC	-	On	On
22	H.261(diff)	CIF	768	56	710.4	INTER+MC	-	On	Off
23	H.261(same)	CIF	768	56	710.4	INTER+MC	-	On	On
24	H.261(diff)	CIF	1536	56	1478.4	INTER+MC	-	On	Off
25	H.261(same)	CIF	1536	56	1478.4	INTER+MC	-	On	Off

* Specified value. Actual frame rate may be determined through measurement.

APPENDIX E - DETAILED TEST PLAN AD HOC GROUP MEMBERS

NAME	REPRESENTING
Al Morton	AT&T Communications (Chair)
Tony Schiano	AT&T Communications
David Hayner	Ameritech Srvcs
Keith Kommeyer	Bell Atlantic
Ron McConnell	BellCore
Dan Wirth	BellCore
Dan Klenke	Compression Labs, Inc
R. Schaphorst	Delta Information Sys
Neil Randall	Delta Information Sys
John Roth	Delta Information Sys
Greg Cermak	GTE Labs
Eric Hauch	Government of Canada
Stephen Wolf	NTIA/ITS.N3
Rich Baker	PictureTel Corp
Robert Reynolds	(formally) PictureTel Corp
Xian-Cheng Yuan	PictureTel Corp
Greg Onyszchuk	Telecom Canada
Doug Stevens	Tektronix
John Grigg	US West
Joe Duran	VTEL

DATA ANALYSIS AD HOC GROUP MEMBERS

NAME	REPRESENTING
Richard Schaphorst	Delta Information Sys (Convener)
Bill Coufal	US West (Secretary)
Al Morton	AT&T Communications
Paul Tukey	BellCore
Dan Wirth	BellCore
C.Frank Taylor	Bell South
Neil Randall	Delta Information Sys
Greg Cermak	GTE Labs
Eric Hauch	Government of Canada
Edwin Crow	NTIA/ITS.N3
Arthur Webster	NTIA/ITS.N3
Stephen Wolf	NTIA/ITS.N3
Rich Baker	PictureTel Corp
Marshall Schachtman	PictureTel Corp

APPENDIX F - RANDOM PRESENTATION ORDERS

PRESENTATION ORDER -- RED TEAM TAPES

HRCs 1 4 7 8 13 15 19 20 22 24

	SESSION 1	SESSION 2	SESSION 3	SESSION 4
	Scene/Type	Scene/Type	Scene/Type	Scene/Type
1	13-f 5-A	19-k 8-A	22-n 9-B	15-s 7-D
2	20-i 8-E	1-s 1-D	4-f 2-A	13-q 5-C
3	4-q 2-C	7-b 3-B	7-u 3-D	7-a 3-B
4	19-x 8-D	19-j 8-A	8-o 4-C	8-q 4-C
5	15-l 7-A	*22-c 9-D	22-t 9-D	19-m 8-B
6	8-r 4-C	19-b 8-B	7-k 3-A	4-s 2-D
7	15-e 7-B	1-k 1-A	4-v 2-D	22-i 9-E
8	1-p 1-C	24-p 9-C	8-y 4-E	8-w 4-B
9	19-n 8-B	4-y 2-E	19-l 8-A	20-s 8-D
10	4-r 2-C	19-t 8-D	8-t 4-D	24-e 9-B
11	24-c 9-D	7-y 3-E	20-w 8-B	15-p 7-C
12	8-e 4-B	4-w 2-B	7-f 3-A	22-v 9-D
13	7-q 3-C	#0-t 1-D	20-n 8-B	15-j 7-A
14	15-m 7-B	13-i 5-E	7-h 3-E	20-y 8-E
15	13-g 5-C	24-l 9-A	4-d 2-C	#0-s 1-D
16	22-k 9-A	1-b 1-B	22-s 9-D	13-l 5-A
17	8-g 4-C	22-u 9-D	4-a 2-B	24-d 9-C
18	15-x 7-D	7-l 3-A	*19-p 8-C	8-m 4-B
19	1-r 1-C	19-d 8-C	24-n 9-B	24-r 9-C
20	7-t 3-D	22-b 9-B	19-o 8-C	13-a 5-B
21	13-w 5-B	4-x 2-D	1-t 1-D	19-h 8-E
22	15-f 7-A	19-r 8-C	8-a 4-B	13-d 5-C
23	1-w 1-B	22-x 9-D	20-f 8-A	19-u 8-D
24	8-v 4-D	8-p 4-C	24-g 9-C	7-i 3-E
25	4-h 2-E	4-j 2-A	8-i 4-E	19-v 8-D
26	22-l 9-A	7-r 3-C	13-c 5-D	24-h 9-E
27	20-x 8-D	*22-c 9-D	7-o 3-C	7-d 3-C
28	22-g 9-C	7-p 3-C	1-v 1-D	*24-x 9-D
29	8-j 4-A	8-c 4-D	8-l 4-A	13-y 5-E
30	22-h 9-E	7-w 3-B	1-m 1-B	7-n 3-B
31	*20-o 8-C	24-y 9-E	15-u 7-D	24-u 9-D
32	13-e 5-B	20-p 8-C	22-e 9-B	7-j 3-A
33	15-i 7-E	4-u 2-D	13-j 5-A	15-h 7-E
34	20-l 8-A	24-a 9-B	20-r 8-C	20-c 8-D
35	15-b 7-B	4-i 2-E	1-h 1-E	24-o 9-C
36	19-y 8-E	8-k 4-A	8-s 4-D	7-e 3-B
37	15-t 7-D	15-q 7-C	1-g 1-C	24-t 9-D
38	1-e 1-B	13-x 5-D	8-u 4-D	20-m 8-B
39	15-o 7-C	4-n 2-B	15-a 7-B	24-j 9-A
40	24-v 9-D	1-q 1-C	8-f 4-A	7-m 3-B
41	1-d 1-C	22-w 9-B	20-g 8-C	24-k 9-A
42	24-b 9-B	19-c 8-D	1-a 1-B	20-a 8-B
43	8-d 4-C	24-f 9-A	7-s 3-D	*24-x 9-D
44	20-u 8-D	1-y 1-E	4-b 2-B	13-o 5-C
45	24-i 9-E	13-t 5-D	*19-p 8-C	4-l 2-A

46	#0-c 1-D	24-q 9-C	1-f 1-A	15-y 7-E
47	19-q 8-C	20-b 8-B	19-s 8-D	19-f 8-A
48	7-v 3-D	1-x 1-D	24-w 9-B	4-m 2-B
49	19-i 8-E	8-n 4-B	15-v 7-D	22-q 9-C
50	*20-o 8-C	15-d 7-C	22-p 9-C	1-j 1-A
51	1-n 1-B	22-a 9-B	13-u 5-D	20-q 8-C
52	4-t 2-D	7-c 3-D	#0-h 1-E	22-m 9-B
53	13-h 5-E	4-p 2-C	8-b 4-B	4-k 2-A
54	20-t 8-D	1-u 1-D	4-g 2-C	15-w 7-B
55	24-m 9-B	4-e 2-B	22-j 9-A	1-o 1-C
56	20-v 8-D	24-s 9-D	8-h 4-E	15-c 7-D
57	13-r 5-C	13-b 5-B	15-r 7-C	1-i 1-E
58	19-e 8-B	19-g 8-C	1-l 1-A	7-x 3-D
59	15-k 7-A	1-c 1-D	13-m 5-B	22-y 9-E
60	4-c 2-D	13-k 5-A	4-o 2-C	20-j 8-A
61	20-e 8-B	22-o 9-C	20-k 8-A	13-s 5-D
62	7-g 3-C	15-n 7-B	13-v 5-D	20-h 8-E
63	8-x 4-D	13-p 5-C	19-w 8-B	22-r 9-C
64	20-d 8-C	22-f 9-A	15-g 7-C	13-n 5-B
65		19-a 8-B		22-d 9-C

PRESENTATION ORDER -- GREEN TEAM TAPES

HRCs 2 5 6 10 14 15 16 17 20 23

	SESSION 1	SESSION 2	SESSION 3	SESSION 4
	Scene/Type	Scene/Type	Scene/Type	Scene/Type
1	2-v 1-D	2-e 1-B	10-t 4-D	6-g 3-C
2	17-k 7-A	17-q 7-C	15-d 7-C	16-h 7-E
3	14-u 6-D	5-x 2-D	20-u 8-D	20-n 8-B
4	15-e 7-B	16-g 7-C	23-a 9-B	5-u 2-D
5	2-h 1-E	#0-c 1-D	*20-q 8-C	14-e 6-B
6	23-e 9-B	15-h 7-E	17-t 7-D	2-j 1-A
7	6-q 3-C	14-x 6-D	14-m 6-B	16-x 7-D
8	17-y 7-E	16-r 7-C	2-r 1-C	20-h 8-E
9	2-p 1-C	23-x 9-D	17-i 7-A	15-f 7-A
10	15-v 7-D	16-k 7-A	10-q 4-C	14-v 6-D
11	14-y 6-E	10-i 4-E	14-a 6-B	10-h 4-E
12	10-w 4-B	14-g 6-C	15-t 7-D	23-d 9-C
13	23-g 9-C	20-m 8-B	6-m 3-B	14-f 6-A
14	5-f 2-A	15-x 7-D	17-s 7-D	20-o 8-C
15	14-h 6-E	14-n 6-B	5-y 2-E	17-j 7-A
16	*20-d 8-C	17-u 7-D	16-p 7-C	#0-h 1-E
17	6-c 3-D	14-j 6-A	23-n 9-B	5-l 2-A
18	16-n 7-B	5-s 2-D	10-g 4-C	*23-s 9-D
19	5-g 2-C	17-n 7-B	6-k 3-A	16-e 7-B
20	6-e 3-B	23-i 9-A	2-g 1-C	10-f 4-A
21	5-o 2-C	20-g 8-C	6-u 3-D	6-w 3-B
22	17-e 7-B	16-a 7-B	20-y 8-E	14-l 6-A
23	20-f 8-A	20-i 8-E	10-r 4-C	20-r 8-C
24	17-o 7-C	16-j 7-A	16-c 7-D	10-n 4-B
25	6-v 3-D	2-b 1-B	6-l 3-A	6-j 3-A
26	20-a 8-B	10-d 4-C	23-p 9-C	23-q 9-C
27	6-p 3-C	6-h 3-E	#0-s 1-D	10-l 4-A
28	20-s 8-D	14-d 6-C	10-o 4-C	15-y 7-E
29	23-k 9-A	20-v 8-D	6-x 3-D	23-w 9-B
30	5-v 2-D	10-a 4-B	2-y 1-E	16-q 7-C
31	2-i 1-E	15-c 7-D	16-u 7-D	10-x 4-D
32	5-p 2-C	5-i 2-E	6-r 3-C	17-r 7-C
33	23-y 9-E	2-k 1-A	23-j 9-A	2-f 1-A
34	20-e 8-B	17-h 7-E	20-c 8-D	5-c 2-D
35	23-l 9-A	23-v 9-D	5-d 2-C	2-m 1-B
36	15-a 7-B	14-i 6-E	2-c 1-D	*23-s 9-D
37	10-v 4-D	17-f 7-A	17-l 7-A	5-n 2-B
38	15-p 7-C	10-c 4-D	*20-q 8-C	16-y 7-E
39	20-k 8-A	16-m 7-B	5-a 2-B	2-w 1-B
40	6-t 3-D	*23-u 9-D	20-l 8-A	16-s 7-D
41	15-k 7-A	10-e 4-B	14-t 6-D	23-f 9-A
42	14-r 6-C	14-s 6-D	6-i 3-E	6-s 3-D
43	10-b 4-B	23-h 9-E	5-k 2-A	14-p 6-C
44	16-d 7-C	6-b 3-B	15-q 7-C	5-m 2-B
45	6-y 3-E	15-l 7-A	14-c 6-D	15-s 7-D
46	17-x 7-D	23-t 9-D	6-o 3-C	14-o 6-C
47	2-n 1-B	15-o 7-C	2-l 1-A	17-m 7-B

48	20-p 8-C	20-b 8-B	15-n 7-B	14-k 6-A
49	16-b 7-B	17-c 7-D	20-j 8-A	20-t 8-D
50	2-u 1-D	14-w 6-B	16-t 7-D	17-a 7-B
51	5-h 2-E	23-o 9-C	6-a 3-B	23-r 9-C
52	23-b 9-B	10-j 4-A	5-t 2-D	10-y 4-E
53	17-d 7-C	16-w 7-B	17-w 7-B	5-j 2-A
54	6-f 3-A	2-d 1-C	10-p 4-C	15-w 7-B
55	16-v 7-D	15-i 7-E	5-e 2-B	10-s 4-D
56	14-q 6-C	5-q 2-C	6-d 3-C	15-j 7-A
57	23-m 9-B	16-i 7-E	16-f 7-A	2-q 1-C
58	*20-d 8-C	2-x 1-D	10-m 4-B	15-b 7-B
59	2-a 1-B	16-o 7-C	2-t 1-D	23-c 9-D
60	5-r 2-C	10-k 4-A	6-n 3-B	17-b 7-B
61	16-l 7-A	15-u 7-D	2-s 1-D	10-u 4-D
62	#0-t 1-D	5-b 2-B	14-b 6-B	5-w 2-B
63	17-p 7-C	15-r 7-C	17-v 7-D	17-g 7-C
64	20-x 8-D	*23-u 9-D	2-o 1-C	20-w 8-B
65		15-m 7-B		15-g 7-C

PRESENTATION ORDER -- ORANGE TEAM TAPES

HRCs 3 4 9 11 12 17 18 20 21 25

	SESSION 1	SESSION 2	SESSION 3	SESSION 4
	Scene/Type	Scene/Type	Scene/Type	Scene/Type
1	4-v 2-D	20-t 8-D	*21-o 8-C	21-b 8-B
2	20-i 8-E	4-g 2-C	3-i 1-E	11-r 5-C
3	3-s 1-D	25-n 9-B	25-c 9-D	21-k 8-A
4	4-y 2-E	17-v 7-D	9-w 4-B	#0-t 1-D
5	9-n 4-B	3-o 1-C	17-j 7-A	17-d 7-C
6	20-y 8-E	4-s 2-D	4-d 2-C	9-s 4-D
7	17-x 7-D	18-w 7-B	21-s 8-D	4-j 2-A
8	12-f 5-A	25-p 9-C	9-l 4-A	12-o 5-C
9	9-q 4-C	20-e 8-B	25-d 9-C	18-v 7-D
10	20-h 8-E	4-q 2-C	11-l 5-A	11-q 5-C
11	25-m 9-B	3-k 1-A	21-a 8-B	20-v 8-D
12	#0-h 1-E	9-h 4-E	11-u 5-D	3-e 1-B
13	18-k 7-A	12-x 5-D	4-k 2-A	12-q 5-C
14	11-m 5-B	20-d 8-C	11-t 5-D	3-m 1-B
15	18-p 7-C	3-u 1-D	3-r 1-C	12-p 5-C
16	3-t 1-D	9-b 4-B	9-c 4-D	3-a 1-B
17	21-g 8-C	25-y 9-E	4-o 2-C	18-g 7-C
18	3-h 1-E	3-b 1-B	20-w 8-B	4-l 2-A
19	9-o 4-C	17-r 7-C	25-u 9-D	9-i 4-E
20	21-w 8-B	25-x 9-D	18-y 7-E	20-o 8-C
21	17-f 7-A	3-g 1-C	25-g 9-C	9-j 4-A
22	3-y 1-E	17-e 7-B	20-u 8-D	18-e 7-B
23	9-a 4-B	12-i 5-E	12-e 5-B	*20-p 8-C
24	11-f 5-A	25-r 9-C	20-k 8-E	17-n 7-B
25	4-c 2-D	20-l 8-A	11-g 5-C	3-c 1-D
26	17-y 7-E	9-r 4-C	21-f 8-A	11-p 5-C
27	11-n 5-B	18-x 7-D	4-a 2-B	20-x 8-D
28	25-f 9-A	9-m 4-B	25-t 9-D	12-a 5-B
29	20-g 8-C	18-h 7-E	4-b 2-B	18-f 7-A
30	11-c 5-D	21-d 8-C	21-u 8-D	25-h 9-E
31	25-l 9-A	25-i 9-E	12-n 5-B	4-e 2-B
32	12-y 5-E	*21-p 8-C	4-i 2-E	9-p 4-C
33	25-j 9-A	17-t 7-D	*21-o 8-C	4-f 2-A
34	3-x 1-D	25-o 9-C	18-u 7-D	9-y 4-E
35	18-o 7-C	17-w 7-B	9-d 4-C	11-b 5-B
36	21-m 8-B	21-y 8-E	11-y 5-E	4-p 2-C
37	12-l 5-A	25-e 9-B	3-p 1-C	18-t 7-D
38	20-r 8-C	21-q 8-C	12-t 5-D	11-j 5-A
39	4-n 2-B	11-w 5-B	21-i 8-E	18-a 7-B
40	*25-v 9-D	9-f 4-A	18-s 7-D	21-c 8-D
41	11-k 5-A	11-a 5-B	12-m 5-B	4-r 2-C
42	9-u 4-D	21-t 8-D	20-j 8-A	17-h 7-E
43	12-j 5-A	9-e 4-B	#0-c 1-D	20-a 8-B
44	17-m 7-B	12-s 5-D	18-b 7-B	17-k 7-A
45	4-x 2-D	20-m 8-B	21-r 8-C	12-b 5-B
46	18-l 7-A	17-p 7-C	9-k 4-A	20-c 8-D
47	11-d 5-C	9-x 4-D	17-a 7-B	25-q 9-C

48	17-l 7-A	21-n 8-B	12-k 5-A	18-i 7-E
49	3-v 1-D	12-g 5-C	20-n 8-B	11-o 5-C
50	20-q 8-C	25-w 9-B	25-k 9-A	20-f 8-A
51	9-t 4-D	18-d 7-C	21-v 8-D	11-h 5-E
52	3-d 1-C	4-w 2-B	9-g 4-C	21-j 8-A
53	12-v 5-D	25-s 9-D	17-c 7-D	3-n 1-B
54	3-q 1-C	11-e 5-B	21-h 8-E	4-u 2-D
55	21-x 8-D	4-h 2-E	3-l 1-A	18-m 7-B
56	11-i 5-E	17-b 7-B	11-s 5-D	12-d 5-C
57	17-s 7-D	11-v 5-D	18-r 7-C	20-s 8-D
58	21-l 8-A	17-g 7-C	3-w 1-B	25-b 9-B
59	18-c 7-D	12-u 5-D	18-q 7-C	12-r 5-C
60	12-w 5-B	17-i 7-E	21-e 8-B	9-v 4-D
61	17-o 7-C	*21-p 8-C	12-c 5-D	17-q 7-C
62	*25-v 9-D	12-h 5-E	18-j 7-A	4-t 2-D
63	18-n 7-B	25-a 9-B	11-x 5-D	*20-p 8-C
64	3-f 1-A	#0-s 1-D	20-b 8-B	3-j 1-A
65		4-m 2-B		17-u 7-D

APPENDIX G - EQUIVALENT METHOD of VIEWER RANDOMIZATION

This appendix presents a method of viewer randomization that is equivalent to that proposed in Section 2.9.2 of the Subjective Test Plan. The new method permits gradual viewer recruiting through sequential assignment to random teams and viewing positions. This method was discussed and accepted by the full Working Group at the November 10, 1993 meeting in San Jose, CA.

At the time this appendix was prepared, it was known that each lab would have all 12 subjective viewing tapes at the same time. Therefore, it is possible to randomize the tape teams (RED, GREEN, and ORANGE) and their respective random orderings (e.g. 1423 2134 3214 etc.) and assign a particular sub-team (e.g. G1 G4 G2 G3) to a viewer as he is recruited. This preserves the balance achieved through randomization and allows the labs to go forward with the testing while the recruiting process continues.

The same random numbers that appear in Table 7 of Section 2.9.2 will be used. They will now be used to order the sub-teams, however, instead of the viewers. The viewers will receive numbers from the Table 7 as they are recruited (or scheduled).

Each viewer will be given a number from Table 7 assigning each to a sub-team and seat position. The assignment will be made according to Table 7 (for each lab) proceeding down the columns and disregarding any numbers beyond 36 (or beyond the expected number of viewers). For example, the first-recruited subject for Lab X will be viewer #8 and the second-recruited subject for Lab X will be viewer #3.

Each sub-team is composed of three viewers and hence will have three numbers associated with it (one for the viewer assigned to the Left seat, one for the Center seat, and one for the Right seat). Within each sub-team of three subjects the first-listed subject will sit in the Left chair in each session, the second-listed subject will sit in the Center chair, and the third-listed subject will sit in the Right chair. For example:

Viewer Numberings for LabX		Viewer Numberings for LabY	
Viewer Number L C R	Subteam (tape order)	Viewer Number L C R	Subteam (tape order)
1,2,3	G1G4G2G3	1,2,3	03010204
4,5,6	G2G1G3G4	4,5,6	04010203
7,8,9	G3G2G1G4	7,8,9	01020304
.		Etc.	
.			
.			
34,35,36	R3R4R1R2		

Accordingly, the first-recruited subject for Lab X (Viewer #8) will be assigned the sub-team G3G2G1G4 and will sit in the Center seat. Similarly, the second-recruited subject (Viewer #3) will be assigned to sub-team G1G4G2G3 and will sit in the Right seat.

In order to facilitate scheduling subjects, substitutions may be allowed within sub-teams of the same "color" (R,G,O), but not across "colors." For example, if the viewer assigned to sub-team/seat number 2 (i.e. viewer #2), and scheduled to view tapes G1G4G2G3, was not able to make the scheduled showing, the lab could fill the vacant slot with any other

viewer who was scheduled to view GREEN tapes. However, no more than three such substitutions will be allowed within any color group.

This simpler and potentially more convenient way to randomize viewers is expected to facilitate completion of the testing in a timely manner.

APPENDIX H - CRITERIA AND THRESHOLDS FOR NORMAL VISION

This appendix gives the criteria used to determine whether or not each candidate viewer possesses normal vision. The test plan requires use of scores from only those viewers determined to have normal visual acuity and color discrimination. The scores of viewers who fail either the acuity or color tests are set aside from the main analysis.

When attempting to determine the normal threshold scores on the acuity tests, it was found that a consensus did not exist among the optometrists consulted. Furthermore, the composition of color vision test materials vary by supplier and the normal vision thresholds vary by specific test. This appendix documents the thresholds (and tests) that were used in association with this test plan.

The table below defines normal vision for the purpose of this test program.

CRITERIA	LAB	THRESHOLD SCORES
Acuity at 6H (seated)	ITS	zero errors on line 7 (20/25 or better)
Acuity at 6H (seated)	GTE	zero errors on line 7 (20/25 or better)
Acuity at 6H seated standing	DIS/NCS	zero errors on line 7 ≤1 error on line 7
Color Discrimination		
Pseudo-Isochromatic Plates for Testing Color Perception	ITS	≥5 correct out of 7 (sub-set of plates: 86, 15, 56, 57, 56, 47, 74)
Pseudo-Isochromatic Plates for Testing Color Perception	GTE	≥15 correct out of 18 (complete set of plates)
Pseudo-Isochromatic Plates for Testing Color Perception	DIS/NCS	≥10 correct out of 14 (complete set of plates)