



STUDY GROUP 9 – CONTRIBUTION 15

SOURCE*: ASSOCIATE RAPPORTEUR FOR MEASUREMENT AND CONTROL OF THE QUALITY OF SERVICE

TITLE: DRAFT NEW ITU-T RECOMMENDATION J.S VQ1 - SUBJECTIVE PICTURE QUALITY ASSESSMENT FOR DIGITAL CABLE TELEVISION SYSTEMS

Summary

This Recommendation specifies subjective picture quality assessment methods for digital cable television applications.

1. Scope

This Recommendation describes the subjective assessment of picture quality for digital cable television systems. It concerns all of the television chain from the signal source to user's receiver. This chain may contain satellite links, terrestrial links and/or cable links. The assessment is made for home receivers assuming a home viewing environment.

2. References

1. ITU-R Recommendation BT 1129-1, 1997, „Subjective assessment of standard definition digital television (SDTV) systems”.
2. ITU-R Recommendation BT 500-7, 1997, „Methodology for the subjective assessment of the quality of television pictures”.
3. ITU-R Recommendation BT1210, 1997, „Test materials to be used in subjective assessment”.
4. ITU-T Recommendation J 84, 1995, „Distribution of digital multi - programme signals for television, sound and data services through SMATV networks”.
5. ITU-T Recommendation J 83, 1995, „Digital multi-programme systems for television sound and data services for cable distribution”.
6. ITU-T Recommendation P 910, 1995, „Subjective video quality assessment methods for multimedia applications”.
7. ITU-T Recommendation P 920, 1995, „Interactive test methods for audio visual communications”.

8. ITU-T Recommendation P 80, 1993, „Methods for subjective determination of transmissions quality”.
9. ITU-T Recommendation 814, 1994, „Specification and alignment procedures for setting of brightness and contrast of display”.
10. ITU-T Recommendation 815, 1994, „Specification of a signal for measurement of contrast ratio and display”.

3. Terms and Definitions

For the purposes of this Recommendation the following definitions apply.

PLUGE: test signal consist of a peak white level path surrounded by four black level patches, all set against a background of gray, which level is 50% of the peak signal.

4. Abbreviations

For the purpose of this Recommendation the following abbreviations are used:

PVD	- Preferred Viewing Distance
SDTV	- Standard Definition Television
PS	- Programme Segment
QP	- Quality Parameter
TS	- Test Session
TP	- Test Presentation
SSCQE	- Single Stimulus Continuous Quality Evaluation.

5. General

Subjective assessment methods are used to establish the performance of television systems using measurements that more directly anticipate the reactions of those who might view the systems tested. In this regard, it is understood that it may not be possible to fully characterize system performance by objective means; consequently, it is necessary to supplement objective measurements with subjective measurements.

In subjective method test pictures after passing through measured line are assessed by a group of assessors in specific viewing conditions and results of these tests are treated statistically.

In ITU - R Recommendation BT. 500-7 two classes of subjective assessment are given, in a few methods are representative of each class.

Introduction of digital coding produces impairments to picture quality which are scene dependent and time varying conventional ITU - R methodologies alone are not sufficient to assess this type of material. Furthermore methods of laboratory testing do not replicate home viewing condition of cable television.

For the evaluation of these types of picture quality new methods are developed and tested. Some of them are given in Annex 1.

All these methods are still under study and need validation.

6. Test material

A number of approaches may be taken in establishing the kinds test material required in television assessments. In practice, however, particular kinds of test materials should be used to address particular assessment problems. A survey of typical assessment problems and of test materials used to address these problems is given in Table 1.

TABLE 1
Selection of test material

Assessment problem	Material used
Overall performance with average material	General, „critical but not unduly so”
Capacity, critical applications (e.g. contribution)	Range, including very critical material for the application tested
Performance of „adaptive” systems	Material very critical for „adaptive” scheme used
Identify weaknesses and possible improvements	Critical, attribute-specific material
Identify factors on which systems are seen to vary	Wide range of very rich material
Conversion among different standards	Critical for differences (e.g. field rate)

Some parameters may give rise to a similar order of impairments for most pictures or sequences. In such cases, results obtained with a very small number of pictures or sequences (e.g. two) may still provide a meaningful evaluation.

However, digital systems frequently have an impact which depends heavily on the scene or sequence content. In such cases, there will be, for the totality of programme hours, a statistical distribution of impairment probability and picture or sequence content. Without knowing the form of distribution, which is usually the case, the selection of test material and interpretation of results must be done very carefully.

In general, it is essential to include critical, because it is possible to take this into account when interpreting results, but it is not possible to extrapolate from non-critical material. In cases where scene or sequence content affects results, the material should be chosen to be „critical but not unduly so” for the system under test. The phrase „not unduly so” implies that the pictures could still conceivably form part of normal programme hours. At least four items should, in such cases, be used: for example, half of which are definitely critical, and which are moderately critical.

It is recommended that at least six picture sequences be used in the assessment, plus an additional one to be used for training purposes prior to the start of the trial. The sequences should range between moderately critical and in the context of the bit-rate reduction application being considered.

In subjective assessment, still pictures and moving sequences may be selected from those listed in Recommendation ITU-R BT. 1210, Annex 1. In this respect, it should be noted that digitally stored pictures and sequences, being the most reproducible source signals, are the preferred sources for assessment.

Annex 1
ASSESSMENT METHOD

1. Single stimulus continuous quality method (SSCQE)

In this method observers are asked to assess instantaneous picture quality continuously during a test session with subjects viewing the material one without a source reference.

1.1 Continuous assessment of overall quality

1.1.1 Recording device and set-up

An electronic recording handset connected to a computer should be used for recording the continuous quality assessment from the subjects. This device should have the following characteristics:

- slider mechanism without any sprung position;
- linear range of travel of 10 cm;
- fixed or desk-mounted position;
- samples recorded twice a second.

1.1.2 General form of the test protocol

Subjects should be presented with test sessions of the following format:

- Programme Segment (PS): a PS corresponds to one programme type (e.g. sport, news, drama) processed according to one of the Quality Parameters (QP) under evaluation (e.g. bit rate); each PS should be at least 5 minutes long;
- Test Session (TS): a TS is a series of one or more different combinations PS/QP without separation and arranged in a pseudo-random order. Each TS contains at least once all the PS and QP but not necessarily all the PS/QP combinations; each TS should be between 30 and 60 minutes duration;
- Test Presentation (TP): a TP represents the full performance of a test. A TP can be divided in Test Sessions to cope with maximum duration requirements and in order to assess the quality over all the PS/QP pairs. If the number of PS/QP pairs limited, a TP can be made of a repetition of the same Test Session to perform the test on a long enough period of time.

For service quality evaluation, audio may be introduced. In this case, selection of the accompanying audio material should be considered at that same level of importance as the selection of video material, prior to the test performance.

The simplest test format would use a single Programme Segment and a single Quality Parameter .

1.1.3 Viewing parameters

Viewing conditions should be those currently specified in Recommendations ITU-R BT. 500 and BT. 1129 for subjective assessment in home viewing environment:

- a) ratio of Luminance of inactive screen to peak Luminance: $\leq 0,02$

- b) display brightness and contrast: set up via PLUGE
- c) maximum observation angle relative to the normal: 30°
- d) monitor input: baseband component input
without digital processing
- e) monitor resolution: to be checked and reported
- f) ratio of viewing distance to picture height: 6 H
- g) screen size for a 4/3 format ratio: From 25'' to 29''⁽¹⁾
- h) screen size for a 4/3 format ratio: From 25'' to 29''⁽¹⁾
- i) monitor standard: SDTV
- j) peak luminance: 200 Cd/m²
- k) environmental Illuminance on the screen: 200 Lux⁽²⁾

⁽¹⁾ This screen size satisfies rules of the „Preferred Viewing Distance” (PVD) for PVD = 6H

⁽²⁾ Incident from the environment falling on the screen should be measured perpendicularly on the screen

The viewing distance and the screen sizes are to be selected in order to satisfy the „Preferred Viewing Distance” (PVD). The PVD (in function of the screen sizes) is shown in the following Table and graph.

Screen diagonal (inch)		Screen height (H) (m)	PVD (H)
4/3 ratio	16/9 ratio		
12	15	0.18	9
15	18	0.23	8
20	24	0.30	7
29	36	0.45	6
60	73	0.91	5
> 100	> 120	> 1.53	3-4

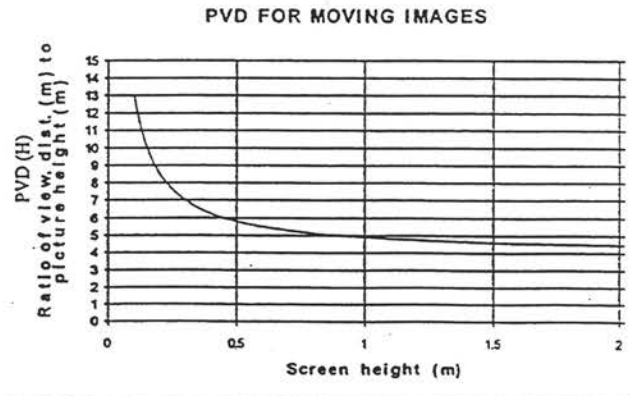


FIGURE 1

This table and this graph are intended to give information on the Preferred Viewing Distance and related screen sizes to be adopted in the Recommendations for specific applications.

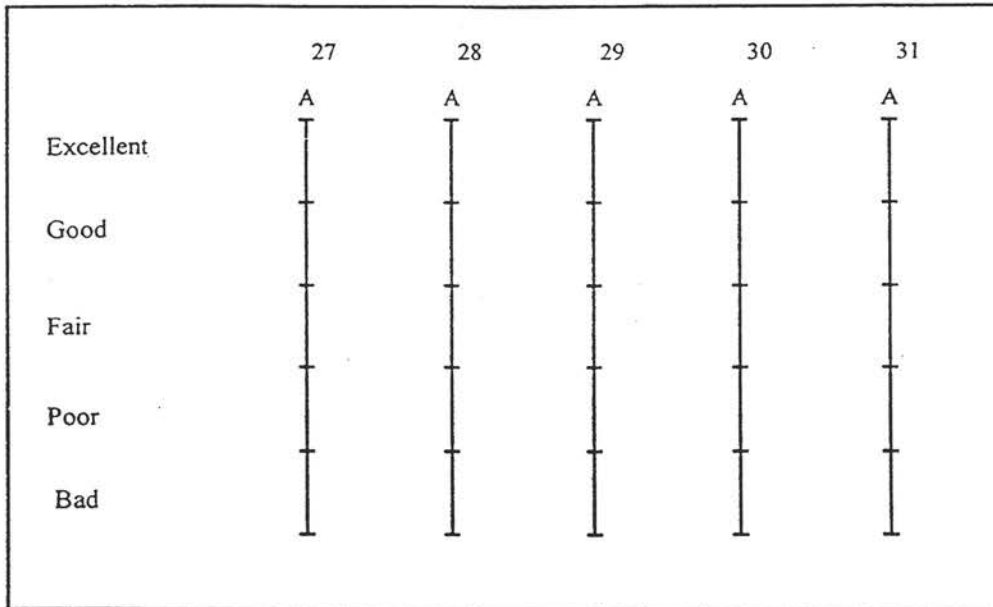
1.1.4 Grading scales

Subjects should be made aware in the test instructions that the range of travel of the handset slider mechanism corresponds to the continuous quality scale.

The observers are simply asked to assess the overall picture quality of each presentation by inserting a mark on a vertical scale. The scales provide a continuous rating system to avoid quantizing errors, but they are divided into five equal lengths which correspond to the normal ITU-R five-point quality

scale. The associated terms categorizing the different levels are the same as those normally used; but here they are included for general guidance and are printed only on the left of the first scale in each row of ten columns on the score sheet. Fig. 2 shows a section of a typical score sheet. Any possibility of confusion between the scale divisions and the test results is avoided by printing the scales in blue and recording the results in black.

FIGURE 2
Portion of quality-rating form using continuous scales



1.1.5 Observers

At least fifteen subjects, non-experts, should be employed.

At least 15 observers should be used. They should be non-expert, in the sense that they are not directly concerned with television picture quality as part of their normal work, and are not experienced assessors*. Prior to a session, the observers should be screened for (corrected-to-) normal visual acuity on the Snellen or Landolt chart, and for normal colour vision using specially selected charts (Ishihara, for instance). The number of assessors needed depends upon the sensitivity and reliability of the test procedure adopted and upon the anticipated size of the effect sought.

1.1.6 Instructions to the observers

In the case of services quality evaluation (with accompanying audio), observers should be instructed to consider the overall quality rather the video quality only.

* Preliminary findings suggest that non-expert observers may yield more critical results with exposure to higher quality transmission and display technologies.

1.1.7 Data presentation, results processing and presentation

Data should be collated from all test session. A single graph of mean quality rating as a function of time, $q(t)$, can therefore be obtained as the mean of all observers' quality gradings per programme segment, quality parameter or per entire test session. (see example in Fig. 3)

This data can be converted to a histogram of probability, $P(q)$, of the occurrence of quality level q . (see example in Fig. 4)

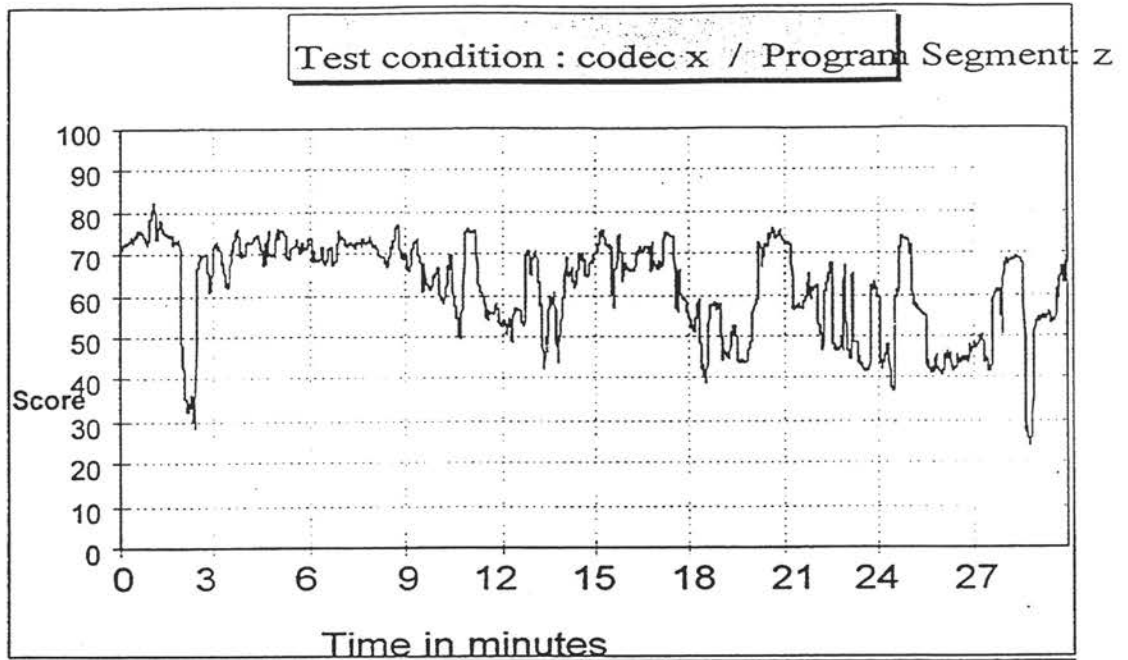


FIGURE 3 - Test Condition: Codec X / Program Segment: Z

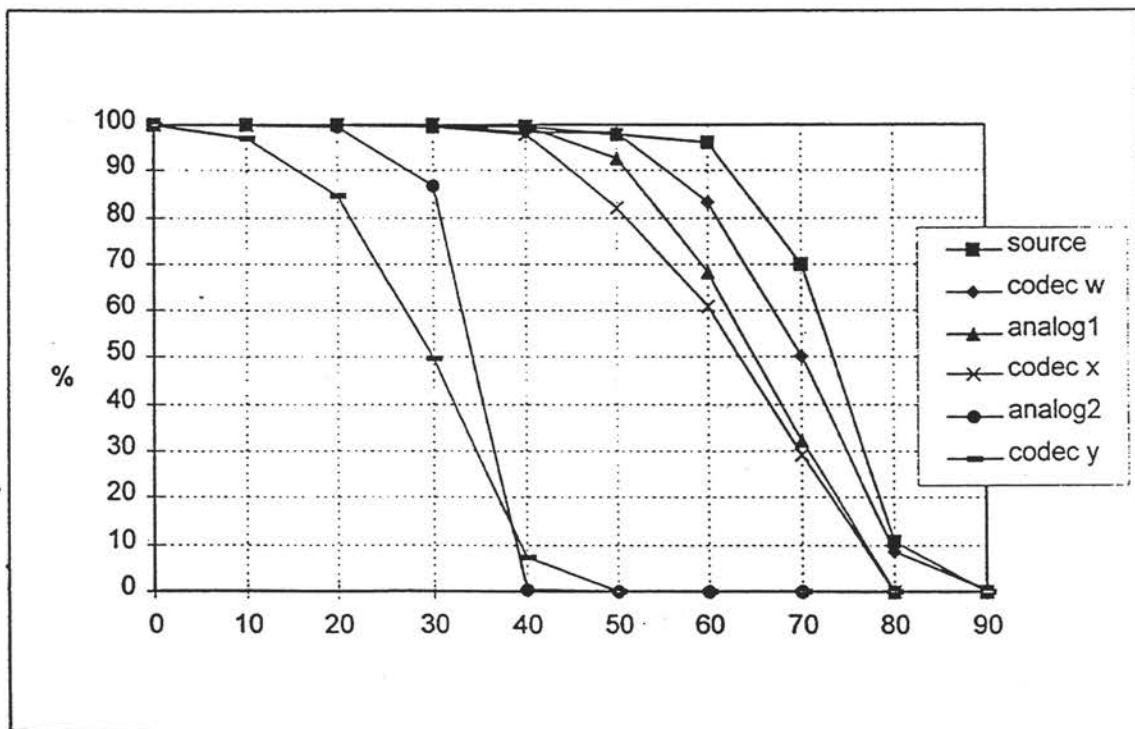


FIGURE 4 - Mean Of Scores Of Voting Sequences On Programmes Segment Z

1.1.8 Calibration of continuous quality results and derivation of a single quality rating

Conventional ITU-R methodologies employed in the past have been able to produce single quality ratings for television sequences. Experiments have been performed which have examined the relationship between the continuous assessment of a coded video sequence, and an overall single quality rating of the same segment. It has already been identified that the human memory effects can distort quality ratings if noticeable impairments occur in approximately the last 10 - 15 s of the sequence. However, it has also been found that this human memory effects could be modelled as a decaying exponential weighting function. Hence a possible third stage in the SSQE methodology would be to process these continuous quality assessments, in order to obtain an equivalent single quality measurement. This is currently under study.
