

RECOMMENDATION ITU-R BT.1788

**Methodology for the subjective assessment of video quality
in multimedia applications**

(Question ITU-R 102/6)

(2007)

Scope

Digital broadcasting systems will permit the delivery of multimedia and data broadcasting applications comprising video, audio, still-picture, text and graphics. This Recommendation specifies non-interactive subjective assessment methods for evaluating the video quality of multimedia applications.

The ITU Radiocommunication Assembly,

considering

- a) that digital broadcasting systems are being introduced in many countries;
- b) that multimedia and data broadcasting services which comprise video, audio, still-picture, text, graphics, etc., have been introduced or are planned to be introduced using digital broadcasting systems;
- c) that multimedia services will involve a broadcasting infrastructure characterized by the possible use of fixed and mobile receivers, fixed and variable frame rates, different image formats, advanced video codecs, packet loss, etc.;
- d) that it will be necessary to specify performance requirements and to verify the suitability of technical solutions considered for each service with the performance requirements of that service;
- e) that such verification will principally involve subjective assessment of video quality under controlled conditions;
- f) that the subjective assessment methodologies specified in Recommendation ITU-R BT.500 may be used for multimedia applications;
- g) that subjective assessment methodologies other than those specified in Recommendation ITU-R BT.500 may also be used;
- h) that the adoption of standardized methods is of importance in the exchange of information between various laboratories,

recommends

- 1 that the general methods of test, i.e. the grading scales and the viewing conditions for the assessment of picture quality described in Annex 1, should be used for laboratory experiments and whenever possible for operational assessments in multimedia applications;
- 2 that the full descriptions of test configurations, test materials, observers and methods should be provided in all test reports;
- 3 that, in order to facilitate the exchange of information between different laboratories, the collected data should be processed in accordance with the statistical techniques detailed in Annex 2.

NOTE 1 – The development of a library of video material appropriate for the subjective assessment of video quality in multimedia applications needs to be further pursued by Radiocommunication Study Group 6.

Annex 1

Description of assessment methods

1 Introduction

Many countries have begun deploying digital broadcasting systems that will permit the delivery of multimedia and data broadcasting applications comprising video, audio, still-picture, text and graphics.

Standardized subjective assessment methods are needed to specify performance requirements and to verify the suitability of technical solutions considered for each application. Subjective methodologies are necessary because they provide measurements that allow industry to more directly anticipate the reactions of end users.

The broadcasting system needed to deliver multimedia applications is markedly different from the one currently in use: information is accessed through fixed and/or mobile receivers; the frame rate can be fixed or variable; the possible image size has a large range (i.e. SQCIF to HDTV); the video is typically associated with embedded audio, text and/or sound; the video may be processed with advanced video codecs; and the preferred viewing distance is highly dependent on the application.

The subjective assessment methods specified in Recommendation ITU-R BT.500 should be applied in this new context. In addition, investigations of multimedia systems might be carried out with new methodologies to meet the user requirements of the characteristics of the multimedia domain.

This Recommendation describes non-interactive subjective assessment methods for evaluating the video quality of multimedia applications. These methods can be applied for different purposes including, but not limited to: selection of algorithms, ranking of audiovisual system performance and evaluation of the video quality level during an audiovisual connection.

Terms and definitions relating to this Recommendation can be found in Appendix 3 to Annex 1.

2 Common features

2.1 Viewing conditions

Recommended viewing conditions are listed in Table 1. The size and the type of display used should be appropriate for the application under investigation. Since several display technologies are to be used in multimedia applications, all relevant information concerning the display (e.g. manufacturer, model and specifications), used in the assessment should be reported.

When PC-based systems are used to present the sequences, the characteristics of the systems (e.g. video display card) should also be reported.

Table 2 shows an example of the data record for the configuration of multimedia system under test.

If the test images are obtained using a specific decoder-player combination, the images must be separated from the proprietary skin to get an anonymous display. This is necessary to ensure that the quality assessment is not influenced by the knowledge of the originating environment.

When the systems assessed in a test use reduced picture format, such as CIF, SIF or QCIF, etc., the sequences should be displayed on a window of the display screen. The colour of the background on the screen should be 50% grey.

TABLE 1

Recommended viewing conditions as used in multimedia quality assessment

Parameter	Setting
Viewing distance ⁽¹⁾	Constrained: 1-8 H Unconstrained: based on viewer's preference
Peak luminance of the screen	70-250 cd/m ²
Ratio of luminance of inactive screen to peak luminance	≤ 0.05
Ratio of the luminance of the screen, when displaying only black level in a completely dark room, to that corresponding to peak white	≤ 0.1
Ratio of luminance of background behind picture monitor to peak luminance of picture ⁽²⁾	≤ 0.2
Chromaticity of background ⁽³⁾	D ₆₅
Background room illumination ⁽²⁾	≤ 20 lux

⁽¹⁾ Viewing distance in general depends on the application.

⁽²⁾ This value indicates a setting allowing maximum detectability of distortions, for some applications higher values are allowed or they are determined by the application.

⁽³⁾ For PC monitors, the chromaticity of background should approximate as much as possible the chromaticity of "white point" of the display.

TABLE 2

Configuration of the multimedia system under test

Parameter	Specification
Type of display	
Display size	
Video display card	
Manufacturer	
Model	
Image information	

2.2 Source signals

The source signal provides the reference picture directly and the input for the system under test. The quality of the source sequences should be as high as possible. As a guideline, the video signal should be recorded in multimedia files using YUV (4:2:2, 4:4:4 formats) or RGB (24 or 32 bits). When the experimenter is interested in comparing results from different laboratories, it is necessary to use a common set of source sequences to eliminate a further source of variation.

2.3 Selection of test materials

The number and type of test scenes are critical for the interpretation of the results of the subjective assessment. Some processes may give rise to a similar magnitude of impairment for most sequences. In such cases, results obtained with a small number of sequences (e.g. two) should provide a meaningful evaluation. However, new systems frequently have an impact that depends heavily on the scene or sequence content. In such cases, the number and type of test scenes should be selected so as to provide a reasonable generalization to normal programming. Furthermore, 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 scene could still conceivably form part of normal television programming content. A useful indication of the complexity of a scene might be provided by its spatial and temporal perceptual characteristics. Measurements of spatial and temporal perceptual characteristics are presented in more detail in Appendix 1 to Annex 1.

2.4 Range of conditions and anchoring

Because most of the assessment methods are sensitive to variations in the range and distribution of conditions seen, judgment sessions should include the full ranges of the factors varied. However, this may be approximated with a more restricted range, by also presenting some conditions that would fall at the extremes of the scales. These may be represented as examples and identified as most extreme (direct anchoring) or distributed throughout the session and not identified as most extreme (indirect anchoring). If possible, a large quality range should be used.

2.5 Observers

The number of observers after screening should be at least 15. They should be non-expert, in the sense that they are not directly concerned with 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 (e.g. Ishihara).

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.

Experimenters should include as many details as possible on the characteristics of their assessment panels to facilitate further investigation of this factor. Suggested data to be provided could include: occupation category (e.g. broadcast organization employee, university student, office worker), gender and age range.

2.6 Instructions for the assessment

Assessors should be carefully introduced to the method of assessment, the types of impairment or quality factors likely to occur, the grading scale, timing, etc. Training sequences demonstrating the range and the type of the impairments to be assessed should be used with scenes other than those used in the test, but of comparable sensitivity.

2.7 Experimental design

It is left to the experimenter to select the experimental design in order to meet specific cost and accuracy objectives. It is preferable to include at least two replications (i.e. repetitions of identical conditions) in the experiment. Replications make it possible to calculate individual reliability and, if necessary, to discard unreliable results from some subjects. In addition, replications ensure that learning effects within a test are to some extent balanced out. A further improvement in the handling of learning effects is obtained by including a few “dummy presentations” at the beginning of each test session. These conditions should be representative of the presentations to be shown later