

RECOMMENDATION ITU-R BT.710-3

**SUBJECTIVE ASSESSMENT FOR IMAGE QUALITY IN
HIGH-DEFINITION TELEVISION**

(Question ITU-R 211/11)

(1990-1992-1994-1997)

The ITU Radiocommunication Assembly,

considering

- a) that a number of administrations and organizations throughout the world are currently evaluating high-definition television (HDTV) systems, and that in many parts of the world HDTV broadcasting is likely to become the primary medium of the next century;
- b) that subjective assessments are a vital element in HDTV system design and selection;
- c) that Recommendation ITU-R BT.500 outlines general subjective assessment methods, many of the methodological details of which are also appropriate in the context of HDTV;
- d) that, nevertheless, it may be thought helpful to make clear the assessment methods and viewing conditions appropriate for HDTV, in the key areas currently under study, by a separate Recommendation,

recommends

- 1 that subjective assessments of image quality of HDTV systems should be made with the viewing conditions given in Annex 1;
- 2 that subjective assessments of the overall quality of an HDTV image delivered by an emission system should be made using a double-stimulus continuous quality-scale method (Recommendation ITU-R BT.500) with the HDTV studio standard as reference;
- 3 that assessments of the failure characteristics of an analogue HDTV emission system should be made using a double-stimulus impairment scale method (Recommendation ITU-R BT.500) with either the image of the HDTV studio or the image of the unimpaired emission as reference;
- 4 that derivation of picture-content failure characteristics of a digital HDTV system should be made using the procedure as specified in Recommendation ITU-R BT.1129;
- 5 that, in the absence of a high-quality reference, the graphic scaling method or the ratio scaling/magnitude estimation method should be considered for assessments of overall quality of the image (before or after processing) provided by an HDTV studio system;
- 6 that, when a high-quality reference is available, the double-stimulus continuous quality method (Recommendation ITU-R BT.500) should be considered for assessments of overall quality of the image (before, or after processing) provided by an HDTV studio system;
- 7 that, when performance over the range of programme content and transmission conditions likely to be encountered in practice is of issue, the description of composite failure characteristics as in Appendix 2 to Annex 1 of Recommendation ITU-R BT.500 be considered;
- 8 that, in the interpretation of the results of particular studies, due note be taken of any real limitations that current technology may impose upon the results of the study (e.g., bounding effects of pick-up or display devices);
- 9 that care must be taken to distinguish the influence of the display format from that of the basic system format (e.g. any up-conversion). Assessments may be performed in order to take account of the different formats if applicable and appropriate.

NOTE 1 – Information on subjective assessment methods for establishing picture quality in an HDTV environment is given in Annex 2.

NOTE 2 – Information on evaluation factors in global HDTV assessment is given in Annex 3.

ANNEX 1

TABLE 1

Viewing conditions for the subjective assessment of HDTV image quality

Condition	Item	Values ⁽¹⁾
a	Ratio of viewing distance to picture height	3
b	Peak luminance on the screen (cd/m ²) ⁽²⁾	150-250
c	Ratio of luminance of inactive tube screen (beams cut off) to peak luminance ⁽³⁾	≤ 0.02
d	Ratio of the luminance of the screen when displaying only black level in a completely dark room, to that corresponding to peak white ⁽⁴⁾	approximately 0.01
e	Ratio of luminance of background behind picture monitor to peak luminance of picture	approximately 0.15
f	Illumination from other sources ⁽⁵⁾	low
g	Chromaticity of background	D_{65}
h	Angle subtended by that part of the background which satisfies the specification above ⁽⁶⁾ . This should be preserved for all observers	53° high \times 83° wide
i	Arrangement of observers	Within $\pm 30^\circ$ horizontally from the centre of the display. The vertical limit is under study
j	Display size ⁽⁷⁾	1.4 m (55 in)

⁽¹⁾ As it may not be possible currently to achieve these conditions fully for tests, alternative values are given on an interim basis. It should be recognized, however, that the results of tests conducted under the interim conditions may not be, in general, comparable with those obtained in situations in which lower presentation objectives apply.

⁽²⁾ Peak luminance on the screen corresponding to the video signal with 100% amplitude. Values ≥ 70 cd/m² should be used until the specified level becomes technically feasible.

⁽³⁾ This item could be influenced by the room illumination, as well as the contrast range of the display.

⁽⁴⁾ Black level corresponds to the video signal with 0% amplitude.

⁽⁵⁾ Room illumination should be set in order to make it possible to satisfy the conditions c and e.

⁽⁶⁾ A minimum of 28° high \times 48° wide is recommended.

⁽⁷⁾ Values ≥ 76.2 cm (30 in) should be used if displays of the specified size are not available.

ANNEX 2

Subjective assessment methods for picture quality in an HDTV environment

1 Introduction

Methods used in subjective tests of conventional television systems are described in Recommendation ITU-R BT.500. The main concepts of assessment methodology apply equally to all forms of television, but the way in which the detailed specifications of the methods for conventional television apply to HDTV requires careful study.

The ITU-R is examining developments in HDTV and to determine what changes, if any, are required to subjective test methods to accommodate these developments. Studies in this regard are not yet complete.

Before proceeding, it is important to stress the following points:

- picture quality is not the only factor which needs to be considered in the selection of standards. Other factors such as system complexity, availability, future possibilities, etc., must be part of the overall equation;
- the results of subjective assessment experiments are not in themselves laws of physics. They offer guidance for a given set of test conditions, and are not absolute facts about a system;
- the conceptual differences between the quality and impairment scale terms currently used are not uniform; but, traditionally, processing of results uses the approximation that they are so. Studies on alternative assessment methods with fewer shortcomings are being made, but interpretation of the results of current methods must take account of the shortcomings;
- the key element in subjective assessments is often the selection of test material. Guidelines call for material which is critical but not unduly so. Deciding what could be critical needs a full understanding of how HDTV systems work.

2 Picture quality evaluations in an HDTV environment

2.1 Areas for picture quality evaluations

2.1.1 Evaluations of HDTV studio formats

There is a need to evaluate:

- basic picture quality,
- picture quality after downstream processing such as colour-matte, slow motion and picture manipulation, and possible conversion to other formats, including film.

2.1.2 Evaluations of conventional studio formats (and film) derived from HDTV studio sources

There is a need to evaluate the adequacy, in terms of picture quality, of conventional studio formats and of film derived from HDTV studio sources.

2.1.3 Evaluations of HDTV emission formats

There is a need to evaluate:

- basic picture quality;
- failure characteristics;
- echo behaviour; and
- susceptibility to interference.

2.1.4 Evaluations for conventional television pictures embedded in HDTV emissions

Some of the HDTV emission formats currently under consideration include an embedded conventional television format ("backwards compatibility"). Thus, there is a need to evaluate, in terms of picture quality, the adequacy of conventional television pictures embedded in HDTV emissions.

2.2 Issues for picture quality evaluations

2.2.1 Evaluation methods

2.2.1.1 Evaluations of picture quality

The five-grade quality terms currently used in subjective assessments are not uniformly spaced conceptually and difficulties have been noted in comparing results obtained in different laboratories, particularly when language translation of terms is required. Further, due to the sensitivity of quality evaluations using conceptual quality terms to the range of conditions used in the test, it is unwise to interpret terms in an absolute fashion or to compare results from tests conducted using different ranges of quality (e.g. HDTV and conventional television).

A seven-grade quality scale has been used successfully to establish the meaning of HDTV quality and such techniques may be useful in future. Further, alternatives to the five-grade quality methods are presented in § 6 of Recommendation ITU-R BT.500. Nevertheless, on balance, it is suggested that the double-stimulus continuous quality method given in Recommendation ITU-R BT.500 generally be used for quality evaluations in an HDTV environment.

2.2.1.2 Evaluations of picture impairments

To an extent, the same problems have been noted for the five-grade impairment scale as for the five-grade quality scale. On balance, however, it is recommended that, when picture quality impairments are to be evaluated, the double-stimulus impairment method given in Recommendation ITU-R BT.500 generally be used.

2.2.2 Viewing conditions for subjective evaluations in an HDTV environment

2.2.2.1 Evaluations of HDTV studio formats

Report ITU-R BT.801 gives picture presentation objectives for HDTV studio formats.

2.2.2.2 Evaluations of conventional studio formats derived from HDTV studio sources

As these evaluations concern TV systems already considered in UIT-R texts, evaluations of conventional studio formats should use the viewing conditions already agreed and presented in Recommendation ITU-R BT.500.

2.2.2.3 Evaluations of HDTV emission formats

It is unclear how well the picture presentation objectives for HDTV studio pictures relate to conditions likely in home viewing. However, subjective evaluations of HDTV emission formats should take account in some way of the higher performance objectives of the HDTV studio.

It is likely that, due to constraints on emission, HDTV emission formats will be unable to fully reproduce the level of picture quality possible in the HDTV studio. However, in recognition of the objective in emission formats to reproduce, as near as possible, the original studio image and in order to preserve consistency of subjective tests throughout the HDTV studio-emission chain, it is suggested that the viewing conditions given in Annex 1 be used equally for tests of HDTV emission formats, and for tests of HDTV studio formats.

2.2.2.4 Evaluations of conventional television pictures embedded in HDTV emissions

As these involve conventional television pictures, the viewing conditions given in Recommendations ITU-R BT.500 and ITU-R BT.1129 apply.

3 Assessment of the picture quality of HDTV studio formats

3.1 Assessment of basic picture quality

At issue here is the picture quality of the HDTV studio format prior to downstream processing. Factors likely to affect basic picture quality include, but are not confined to, spatial resolution, temporal resolution, colour-gamut, and linearity characteristics. Annex 3 of this Recommendation summarizes work on evaluation factors for assessing HDTV picture quality.

There is general agreement that an increase in colour-gamut and the inclusion of constant luminance coding are desirable goals for the HDTV studio system. However, these features carry with them the need for more complex signal processing at the camera and display, and it may become necessary to evaluate trade-offs between the benefits of these goals and the possible disadvantages due to the complex signal processing.

Evaluation of the value of increased colour-gamut and the impact of additional processing requires the availability of a display having a significantly larger gamut than current cathode ray tube (CRT) displays and a source signal properly processed for that large gamut display. In addition, a current CRT-type display is required with the non-linear processing to transform the large gamut source signal into an appropriate signal for this smaller gamut display. Still pictures containing a range of normal colours plus a few colours that lie outside the smaller gamut should be evaluated by comparing the two displays. One source for high purity colours are balls of yarn containing saturated colours. Materials of this sort can provide very saturated colours that lie outside the gamut of current CRT displays and still lend themselves to reasonable scene composition. Subjective evaluation of such a scene on normal CRT displays and on high purity displays should provide an estimate of the quality gain.

The evaluation of constant-luminance coding methods in comparison to non-constant luminance coding methods should be carried out by comparing a full bandwidth *RGB* display to a display on which either constant-luminance or non-constant-luminance signals can be displayed. The scene subject matter should include detail in saturated colours along with normal scene elements. Shadows formed on balls of saturated red, green and blue yarns are one means for providing the detail in saturated colours.

The methods normally used to assess picture quality (i.e., double-stimulus methods) typically require a reference condition that provides quality superior to that of the system under test. The high quality of an HDTV studio system, however, makes it difficult to find appropriate reference conditions. For this reason, it may be appropriate to use directly-viewed scenes (still and moving) to provide the reference condition for assessments of HDTV studio systems.

3.1.1 Methodology

The double-stimulus continuous quality-scale method could be used. The reference for picture quality assessments could be the scene viewed directly (subject to appropriate framing). The test could be the same scene viewed via the system under test.

3.1.2 Viewing conditions

See Annex 1.

3.1.3 Assessment material

The test material could comprise a number of still pictures and moving sequences. Sources for the still pictures could be either transparencies (rear-illuminated) or photographic prints (directly-illuminated). Sources for the moving sequences could be motion dioramas. The reference condition would be provided when a source is viewed directly, while the test condition would be provided when the same source is viewed via a camera and monitor. Identical framing for the two conditions could be maintained by reflecting the test materials for both on to the same 16:9 viewing mirror. Switching between conditions could be done by shutters in the optical paths. Switching is to be done under experimenter control.

The tests involve implied comparisons of test material from a video camera with the same material viewed directly. To minimize possible contamination of the results by differences implicit to television vs. the “real world”, it will be necessary to control a number of factors. These include:

- *parallax differences*: while viewing, the observer should not be able to move appreciably as this would result in a degree of motion parallax in the directly viewed scene but not in the scene shown on the monitor;
- *visible depth*: the viewing mirror will display alternately the television image and the source scene. The composition and lighting of source scenes should be set to ensure that differences in depth between the television image and the directly-viewed scene are minimized;
- *scene lighting*: the viewing mirror will display alternately the television image and the source scene. The lighting in the source scene will have to be adjusted when the display path is changed to hold intensity and colour temperature (D_{65}) constant in both of the images. The colour temperature may have to be set scene-by-scene.

A number of criteria for the composition of source scenes have been suggested. These include:

- static spatial resolution,
- dynamic spatial resolution,
- luminance rendition,
- colour rendition,
- motion rendition.

In addition, it might be useful to supplement these with other, special-purpose scenes. These might assess:

- apparent depth effects (e.g., in panoramic scenes),
- rendition of familiar tones (e.g., skin tones),
- feeling of presence (e.g., in a rapid pan),
- flicker performance (e.g., with large, white sub-fields).

It is important to establish the standard set of test materials to be used for various subjective assessments of HDTV picture quality, as has been done for 4:2:2 materials (see also Recommendation ITU-R BT.1210).

3.1.4 Interpretation of results

The system tested should approximate, as closely as possible, the level of quality provided by the directly-viewed reference. In considering the results, two issues should be kept in mind:

- an HDTV studio system is likely to make compromises among the various features that relate to quality. In addition to considering quality averaged over the various pieces of test material, it would be wise to examine reactions to the individual source scenes in order to identify features that could be improved;
- in interpreting results, it is necessary to identify and, to the extent possible, adjust for possible contamination of the results by technical maturity (state of implementation).

3.2 The assessment of HDTV picture quality following downstream processing

Two areas are considered: post-production processing and standards conversion.

3.2.1 Post-production processing

The major areas of post-production processing are colour-matte, slow-motion and picture manipulation. Assessments, made at the time the Recommendation ITU-R BT.601 4:2:2 standards were developed, suggested that colour-matte is the most demanding post-production operation. For a given field-rate and scan system, this is likely to apply to HDTV.

3.2.1.1 Colour-matte assessments

3.2.1.1.1 Methodology

The double-stimulus impairment scale method should be used provided a full range of picture quality is available. The reference for colour-matte assessments could be a matted picture, using a full-bandwidth *RGB* signal as a foreground. The test could be a matted picture using the reduced colour-difference bandwidth signal as foreground. The matted test and reference pictures should be optimized for quality on a shot-by-shot basis, as this would be the situation in practice. The methodology appropriate, if a full range of picture quality cannot be provided, is still being considered.

3.2.1.1.2 Viewing conditions

See Annex 1.

3.2.1.1.3 Assessment material

The test material should be critical for the types of impairment likely for colour-matte processing. The material which is likely to be most demanding would contain moving fine detail. No specific test sequences for colour-matting in HDTV are known to be available, but moving combs, twisted ribbons, and glass (transparent) may be appropriate for colour-matte evaluations. One still picture is however available. Colour-matte performance depends highly on scene lighting, and care must be taken to ensure this is optimized and consistent (see also Recommendation ITU-R BT.1210).

3.2.1.1.4 Interpretation of results

The test material should not be appreciably impaired relative to the reference material.

3.2.1.2 Slow-motion assessments and picture manipulation assessments

3.2.1.2.1 Methodology, viewing conditions, assessment material, interpretation of results

The assessments in this category pose problems in that a high-quality reference signal is not likely to be available. It is the inclusion of a reference signal which gives the double-stimulus methods their properties. A ratio scaling method is being studied which may be adequately stable and reproducible without a reference. Alternatively, there may, in some cases, be a means of generating high-quality reference sequences. For example, high-quality slow-motion may be possible by a separate shooting of the source sequence at a higher picture rate.

3.2.2 Picture quality following HDTV-to-HDTV standards conversion

3.2.2.1 Methodology

The declared objective of all administrations is to achieve a single worldwide HDTV studio standard, and one of the reasons is to permit international programme exchange without standards conversions. Situations, no doubt, will arise, however, where conversion from other HDTV formats or film will be required. In addition, similar conversions might be needed prior to the generation of an emission format with a different field-rate to the source. In such a case, an investigation of an emission format should consider these types of conversions.

Field-rate standards conversion can give rise to transient temporal artefacts, and in order to provide a better overall system evaluation, a two-tier assessment method is proposed in this case.

3.2.2.1.1 Primary assessments

These are considered to be the main and most useful assessments. A double-stimulus continuous quality scale method should be used. The reference signal should ideally be the same picture or sequence used as an input to the standards converter, but shot using the scanning parameters of the converter output signal. If this is not possible, or if such tests were of interest for other reasons, the reference signal should be the input signal to the converter.

3.2.2.1.2 Auxiliary assessments

A number of expert viewers should be asked, using the single stimulus method (see Recommendation ITU-R BT.500) to assign an overall quality grade to several representative converted programmes. It may also be possible to assess the frequency of detection of artefacts, but this requires further study.

3.2.2.2 Viewing conditions

As given in Annex 1.

3.2.2.3 Test material

3.2.2.3.1 Primary assessments

A relatively large number of still pictures and moving pictures should be used. Possible candidates for HDTV-HDTV assessments are those listed in § 3.1.3 (see also Recommendation ITU-R BT.1210).

3.2.2.3.2 Auxiliary assessment

A number of expert viewers may be asked to scale the overall quality of several programmes of 5-20 min duration, which include examples of different types of movement and scenes with high detail.

3.2.2.3.3 Characteristics of test material

Critical test material for standards conversion is likely to include areas of high detail which have different movement speeds and directions.

3.2.2.4 Interpretation of results

Care must be taken in the interpretation of results, that any inherent quality differences in the two HDTV studio standards are not attributed to the conversion process. The use of reference sequences shot directly in the output standard would assist this.

The subjective quality of the converted picture should be “virtually equivalent” to the input picture, unless it is limited by the parameters of either standard.

4 Assessments of the quality of conventional studio pictures derived from HDTV pictures in the studio environment

4.1 Areas for assessments

The interface between HDTV and conventional TV may imply conversions of line number, frame rate and aspect ratio, although cases without frame rate conversion are also possible. The quality of the conventional picture should be the same as for direct production in the conventional standard.

4.2 Impairments by standards conversion

4.2.1 Impairments due to line number conversions

Conversions involving changes in the number of active lines may result in perceptible disturbances at edges moving vertically. These disturbances may be more pronounced for conversions which increase the number of lines than for those which decrease the number of lines.

4.2.2 Impairments due to frame rate conversions

Conversions involving changes in frame rate will introduce artifacts, such as judder, confined to the moving areas of the picture. The level of these impairments is related to the ratio of the frame rates involved and to the complexity of the conversion algorithm. Some techniques, such as motion adaptive compensation, can reduce these artifacts to very low levels.

4.2.3 Impairments due to aspect ratio conversions

Conversions from the wider aspect ratio of HDTV to the 4:3 aspect ratio of conventional formats may result in the loss of significant picture content or resolution. This is not an area, however, in which subjective evaluation is likely to provide useful guidance.

4.3 Assessment of the quality of conventional quality television derived from an HDTV signal

4.3.1 Methodology

It is evident that the performance of HDTV to conventional television converters can only be completely assessed with moving picture sequences. For the evaluation of the small impairments of the limited range to be expected, the use of the double-stimulus, continuous-quality-scale method is thought to be most useful. The assessors should be asked to view a pair of sequences, one direct in the conventional studio format and the other in the appropriate conventional format but derived from HDTV.

4.3.2 Viewing conditions

The viewing conditions should be as in Recommendations ITU-R BT.500 and ITU-R BT.1129.

4.3.3 Assessment material

A wide range of relatively critical programme material should be used as assessment material (see Recommendation ITU-R BT.1210). The use of still pictures may also be appropriate.

Two other kinds of sequences, likely to be more critical, could be included as well:

- scenes with zoom motion,
- scenes with movements in contrary directions like a market place.

Tests should also be conducted with downstream processed material.

4.3.4 Interpretation

Ideally, an HDTV to conventional television interface should yield the same quality as conventional television direct. Because this demand will probably not be met completely for motion portrayal, the frequency of the assessed picture degradations in television programmes should be investigated. This may imply a two-tier approach as in HDTV-HDTV conversion.

5 Assessment of the quality of HDTV emission systems derived from an HDTV studio standard

5.1 Areas for assessment

The system characteristics which are of interest are as follows:

5.1.1 Basic quality

This is the picture quality under perfect reception conditions, i.e. when the signal-to-noise, S/N , or carrier-to-noise, C/N , ratios are high.

5.1.2 Failure characteristics

This is the relationship between picture quality and noise (which has a characteristic appropriate to the modulation system to be used). The range over which assessments should be made needs to be reviewed in the light of a preliminary run, and should be arranged to give 5-8 points covering the scale range. The range of interest for AM systems is usually an S/N of 25-55 dB, and for FM systems a C/N of 0-30 dB.

5.1.3 Echo behaviour

This is the relationship between picture quality and echo amplitude and delay. It is usually more relevant to AM systems. The range over which assessments need to be made should be reviewed in the light of preliminary runs, but a suitable approach might be to obtain information on three curves, having a delayed signal added to an undelayed signal with 150 ns, 1 μ s, 5 μ s delay respectively, and with echo amplitudes from -5 to -25 dB, compared to the wanted signal.

5.1.4 Interference behaviour

Co-channel and adjacent-channel interference characteristics need to be assessed.

It may well be appropriate to assess § 5.1.2 to 5.1.4 both with and without scrambling.

5.2 Methodology

5.2.1 Basic quality

The basic design problem in HDTV emission is to meet, as near as possible, the visual requirements for HDTV within the bandwidth available. To do this, either or both of spatial and temporal sub-sampling may be used.

Such techniques may introduce detectable impairments, or losses in quality, beyond those attributable to the studio format. Spatial sub-sampling may result in detectable losses in one or more of horizontal, vertical or diagonal resolution. Temporal sub-sampling may result in detectable reductions in the quality of motion portrayal. Spatio-temporal sub-sampling may result in detectable losses in spatial resolution for moving picture sequences.

Clearly, high resolution pictures and moving picture sequences are needed to evaluate HDTV emission formats. However, in order to provide an adequate and representative overall evaluation, a two-tier assessment method is proposed here for basic quality.

5.2.1.1 Primary assessments

These are considered to be the main and most useful assessments. A double-stimulus continuous-quality-scale method should be used. The reference should be the studio source signal and the test signal should be the emission signal.

5.2.1.2 Auxiliary assessments

A number of expert viewers may be asked to scale the overall quality associated with several representative programmes in the emission format. It may also be possible to assess the frequency of detection of artifacts, but this requires further study.

5.2.2 Failure characteristics, echo behaviour and interference behaviour

A double-stimulus impairment scale method should be used following Recommendation ITU-R BT.500.

Two approaches can be taken:

- *cumulative failure characteristics*: for these, the points at which objectionable losses occur relative to the unimpaired high-quality reference are considered;
- *non-cumulative failure characteristics*: for these, the points at which objectionable losses occur relative to the unimpaired emission format are considered.

5.3 Viewing conditions

See Annex 1.

5.4 Test material

5.4.1 Basic quality

The test material should be chosen from a range of high resolution still pictures and moving picture sequences which are critical but not unduly so.

Material which is likely to be critical would require high detail with simultaneous movement at varying speeds and varying directions (see Recommendation ITU-R BT.1210).

5.4.2 Failure characteristics, echo behaviour and interference behaviour

Adequate results should be achieved by using only a small range of still and moving pictures. The overall mean grade can usually meaningfully be calculated and used (see Recommendation ITU-R BT.1210).

5.5 Interpretation of results

5.5.1 Basic quality

It seems reasonable to argue that to be effective, the quality of the HDTV emission signal must be closer to the HDTV studio quality than to the *RGB* quality of conventional television.

As a generality, and for most material, there must be sufficient additional quality, compared to a conventional television. Further, any temporal artifacts must be sufficiently unobtrusive to not detract from the HDTV quality.

5.5.2 Failure characteristics, echo behaviour and interference behaviour

Subject to further study.

6 Assessment of the quality of compatible pictures embedded in HDTV emission formats

6.1 Areas for assessment

Some HDTV emission systems are intended to allow simultaneous reception on HDTV and conventional receivers. Section 5 concerns the assessment of the HDTV emission quality itself. This section concerns the quality of the simultaneously received conventional signal.

In general, a design trade-off must be made between the quality achieved on the HDTV display and the quality achieved on the conventional display, leading to a degree of compatibility based on the level of impairment introduced. This may imply an investigation of the same factors as are listed in § 5, but this time for the compatible picture.

The proposed HDTV emission systems involve temporal processing, and other mechanisms, which might cause impairments to the compatible pictures.

6.2 Methodology

For basic quality, the double-stimulus, continuous-quality method might be used with material prepared directly in the conventional emission format and/or material converted directly from the HDTV studio format as reference. For failure characteristics, echo behaviour, and interference, the double-stimulus, impairment method might be used, with material prepared directly in the conventional emission format (but not otherwise impaired) and/or material converted directly from the HDTV studio format (but not otherwise impaired) as reference. In all cases, the test signal should be the compatibly received picture.

6.3 Viewing conditions

As given in Recommendations ITU-R BT.500 and ITU-R BT.1129 for conventional television.

6.4 Test material

A range of still and moving pictures should be used.

Characteristics of test material should be generally as given for assessments in § 5 (i.e., critical, but not unduly so) (see also Recommendation ITU-R BT.1210).

6.5 Interpretation of results

Interpreting what the quality of “compatible” pictures should be in quantitative terms presents problems, not the least because of the non-integral nature of the scales.

Results for each test picture or sequence should be presented separately.

The quality of the embedded picture should, in principle, be “equivalent” to that of the reference signal. In practice, an agreed “degree of compatibility” must be achieved.

7 Assessment of the quality of motion-picture film derived from HDTV source material

To be investigated.

8 Comparisons of candidate HDTV formats

On occasions, it may be necessary to compare candidate HDTV formats for purposes of selection. It is considered that such comparisons can be used most advantageously to identify the best features of the various formats under test.

8.1 Comparisons of HDTV studio systems

Three ways have been identified in which candidate studio formats can be compared:

- directly, by side-by-side comparison;
- indirectly, by implied comparisons to a common reference condition in a single experiment, and
- theoretically, by establishing relative placements in terms of psychophysically determined optima.

8.2 Direct comparisons

To be investigated.

8.3 Indirect comparisons

An indirect comparison requires a common reference condition with which each system under test is evaluated. The subjective methods normally used for indirect comparisons (i.e., the double-stimulus methods) use reference conditions that, typically, provide quality superior to that of any of the conditions under test.

However, the high quality of candidate HDTV studio systems makes it difficult to find such reference conditions. For this reason, it may be appropriate to use directly-viewed scenes to provide the reference condition.

For a valid indirect test, the directly-viewed reference must be held constant across all systems tested. For still pictures, of course, this can be accomplished by means of transparencies or photographs. For moving images, however, it is necessary to use fully reproducible motion sequences for the reference. This may be done using mechanically controlled scenes (e.g. dioramas).

It is equally important to ensure that, except for differences implicit to the formats themselves, the test materials are held constant across all systems under test. This would be accomplished if the video camera for the system under test is used to capture the reference still or sequence, as long as the reference is held constant.

It should be noted that all systems under consideration should be tested in a single experimental context (i.e. that viewers should see, over the course of the experiment, a random sequence of the systems under test). This may be done by alternating the cameras used to reflect the systems under test. The monitor, which should be held constant, should be selected to be adequate for all systems under test. It might not always be possible to extrapolate the general data applicable to the conditions set up by Annex 1 from the results obtained with the viewing conditions allowed by present equipment. Care should be taken in interpreting the results of the test to distinguish system-standard-related values from those relevant to the practical implementation.

Directly-viewed scenes may provide a reference whose quality is considerably superior to that of the systems under test. This may lead to two issues:

- Differences in subjective reactions to the systems under test may be minimized artificially. When viewers judge, they tend to be influenced by the range and distribution of quality seen. When quality (including that shown by the reference) spans a wide range, cases somewhat similar in quality tend to be judged more similar than they would be if evaluated in a more constrained context or compared directly.

- The preferred test method may change. If conditions (including reference and test) cover a wide range of quality, the double-stimulus, impairment method may be used for direct comparison. However, if conditions span a smaller quality range, the double-stimulus, continuous-quality method is preferred.

Thus, depending upon the purpose of the test, two options arise. If tests are intended to place systems in relation to a “perfect” standard, they may use a superior reference and the double-stimulus, impairment method. In this case, however, fine differences among systems may not be detected. On the other hand, if tests are to make fine discriminations among systems, a superior reference should be avoided and the double-stimulus, continuous-quality method used. In the latter case, it may be necessary to limit the quality of the directly-viewed scene by means of composition, lighting, optical filtering, etc.

8.3.1 Methodology

Depending upon the quality range involved in the test, either the double-stimulus, continuous-quality method or the double-stimulus, impairment method could be used.

If the double-stimulus, continuous-quality method is used, it may be appropriate to consider a variant of this method. In this variant, relatively lengthy exposures are used to encourage the detection of subtle effects, particularly in moving sequences.

If the double-stimulus, continuous-quality method is used, each trial will involve multiple, alternating displays of the reference and test conditions. For half the trials (randomly determined), the reference condition is to be presented first; for the remaining trials, the test condition is to be presented first. If the double-stimulus, impairment method is used, each trial will involve a single alternation between reference and test, with the reference presented first.

The test materials and precautions to be taken to minimize possible contamination of results are as given in § 3.1.3.

Each viewer should see the viewing mirror through a viewing aperture that permits binocular viewing, but little or no head movement. Viewing can take place individually or in small groups. However, if more than one viewer is used at a time, the angle of view (to the scene) must be held constant for all viewers.

As different linguistic groups are known to use quality and impairment scaling terms differently, all tests should be done in a single language with observers fluent in that language.

8.3.2 Viewing conditions

See Annex 1.

8.3.3 Assessment materials

Availability under investigation.

8.3.4 Interpretation of results

Interpretation of results is made on the basis of relative placements of candidate systems relative to the common directly-viewed reference. The issues noted in § 3.1.4 should be kept in mind.

8.4 Theoretical comparisons

The basis of this approach is to consider, parameter by parameter, the placements of candidate systems in terms of the relevant psychophysical ideals. This approach is proposed.

9 Comparisons of candidate HDTV emission formats

As with HDTV studio systems, comparisons may be direct, indirect, or theoretical. Here, only indirect comparisons are considered.

9.1 Basic quality

This generally is as for HDTV studio formats (see § 8.3). Here the double-stimulus, continuous-quality method may be used with a single high-quality reference.

9.2 Failure characteristics

These tests generally are as given in § 5. However, the intent is to compare failure characteristics of all candidate systems.

10 Further issues

Related issues are considered: evaluation methods for conversions from HDTV to 35 mm cine film, the use of the ITU-R quality scale descriptors, interpretations of quality targets in terms of numerical results of assessment, subjective quality and signal-to-noise ratio characteristics of HDTV signals, relationships between picture and sound aspects of HDTV.

Results of a campaign of subjective tests made in Italy, aiming at assessing the preferred viewing distance for HDTV programmes over different types of material and screen sizes, revealed an average preferred viewing distance of about five to six times the height of the picture (5-6 H).

Evidence from earlier studies on conventional resolution television, showed that the preferred viewing distance in this case is 8-9 H .

The relationship between average preferred viewing distance and viewing distance for subjective assessments needs further study.

ANNEX 3

Evaluation factors appropriate to global HDTV assessment

In a recent large-scale study in Canada and the United States of America, viewers evaluated HDTV (MUSE-E via satellite) both in absolute terms and in comparison to studio quality NTSC. Evaluations considered both overall picture quality and specific evaluation factors, including image sharpness, colour quality, motion portrayal, depth portrayal, image brightness, screen size, and screen shape (aspect ratio). The results showed that:

- absolute judgements of HDTV alone concentrate at the end of the quality scale, suggesting possible problems if the results of separate tests with different HDTV systems were to be compared on the basis of absolute quality judgements;
 - viewers were able to respond differentially to the different evaluation factors, suggesting that the specific factors approach may be useful in future evaluations;
 - judgements of overall picture quality were strongly related to most, but not all, of the evaluation factors on which HDTV was perceived to differ from NTSC, suggesting that overall picture quality may fail to fully capture viewer reactions;
 - judgements on specific factors were, to an extent, related, suggesting possible hierarchies amongst the factors used in the evaluations and the possible existence of lower order basic quality factors;
 - judgements of overall picture quality were, to an extent, affected by different evaluation factors, as a function of viewing distance, suggesting a need for careful consideration of the viewing distances to be used in evaluations.
-