International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

J.245

(08/2008)

SERIES J: CABLE NETWORKS AND TRANSMISSION OF TELEVISION, SOUND PROGRAMME AND OTHER MULTIMEDIA SIGNALS

Measurement of the quality of service

Subjective assessment method of picture quality assessed in the home

Recommendation ITU-T J.245

Recommendation ITU-T J.245

Subjective assessment method of picture quality assessed in the home

Summary

Recommendation ITU-T J.245 specifies a subjective test method of picture quality assessed in the home. Picture quality is observed at user-end in home environment on computer monitor or TV set, so only single stimulus method with subjects viewing the material once without the source reference is useful in this case. Details of single stimulus continuous quality evaluation (SSCQE) method especially adapted to observation picture quality in the home on computer or TV set screen are given in this Recommendation.

Source

Recommendation ITU-T J.245 was approved on 13 August 2008 by ITU-T Study Group 9 (2005-2008) under Recommendation ITU-T A.8 procedure.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had not received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database at http://www.itu.int/ITU-T/ipr/.

CONTENTS

Recommendation ITU-T J.245

Subjective assessment method of picture quality assessed in the home

1 Scope

This Recommendation specifies a subjective test method of picture quality assessed in the home. Picture quality is observed at user-end in home environment on computer monitor, or TV set (after set-top-box if needed) so only single stimulus method with subjects viewing the material once without the source reference is useful in this case. Details of SSCQE method especially adapted to observation quality in the home on computer screen or TV set are given in this Recommendation.

Laboratory subjective video quality testing seeks to separate video quality from all other variables (e.g., by using the professional quality monitor, the impact of monitor is removed from subjective scores). By contrast scores from this methodology will have several variables which cannot be separated from the subjective scores (e.g., impact of consumer quality monitor, potentially distracting wall decorations in home viewing environment). This subjective methodology is intended to more precisely analyse the viewing experience of a subject in its own home, and asks different questions than the laboratory tests.

This Recommendation is intended, in particular, for network designers and operators. It gives them guidelines for network design taking into account technical parameters as well as the number of sustainable streams and packet rate available for video services as well as the methodology of subjective picture quality evaluation. It may be also very helpful to service providers to quantify users' complaints about picture quality.

The method described in this Recommendation is not intended to replace a subjective experiment conducted in a controlled test environment (e.g., subjective laboratory) nor regulate consumer equipments.

NOTE 1 – The method in this Recommendation is experimental and has not yet been validated and adequate results have not been reviewed. Therefore, we caution the users to be careful in their use of this method. It should be also noted that the test result is not appropriate to be compared among different home test environments.

NOTE 2 – The structure and content of this Recommendation have been organized for ease of use by those familiar with the original source material; as such, the usual style of ITU-T Recommendations has not been applied.

2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

[ITU-T J.241] Recommendation ITU-T J.241 (2005), *Quality of service ranking and measurement methods for digital video services delivered over broadband IP networks*.

[ITU-T Y.1540] Recommendation ITU-T Y.1540 (2002), *Internet protocol data communication service – IP packet transfer and availability performance parameters*.

[ITU-T Y.1541] Recommendation ITU-T Y.1541 (2006), *Network performance objectives for IP-based services*.

[ITU-R BT.500-11] Recommendation ITU-R BT.500-11 (2002), *Methodology for the subjective assessment of the quality of television pictures*.

[ITU-R BT.1129-2] Recommendation ITU-R BT.1129-2 (1998), *Subjective assessment of standard definition digital television (SDTV) systems*.

[ITU-R BT.1210-3] Recommendation ITU-R BT.1210-3 (2004), *Test materials to be used in subjective assessment*.

[ITU-R BT.1720] Recommendation ITU-R BT.1720 (2005), *Internet protocol data communication service – IP packet transfer and availability performance parameters*.

3 Definitions

This Recommendation does not define any new terms.

4 Abbreviations

This Recommendation uses the following abbreviations:

IP Internet Protocol

PS Program Segment

QP Quality Parameter

SSCQE Single Stimulus Continuous Quality Evaluation

TP Test Presentation

TS Test Session

5 Quality of Service ranking and measurement methods for digital video services in the home

5.1 Background

Digital transport streams based on MPEG-2 encoding have become the prevailing technology for augmenting the experience of digital television services, since it allows combining the distribution of high quality digital television services with the opportunity for end users to enjoy real-time interaction with multimedia service platforms.

This Recommendation is dealing with subjective assessment of picture quality delivered to the home by air, cable, satellite as well as over IP networks. Special attention is given on the latest one because it is a new service and assessing in the home is one possible method for service evaluation.

The broadband fixed communication networks start to be extensively deployed in several countries thus clear opportunities emerge for extending this offer through transport based on IP protocols.

The native shared-access and bidirectional capabilities of an IP network, in fact, offer an ideal environment for providing customers with full end-user interactivity and support for advanced services; this offers advantages over traditional video streaming services. IP-based broadband communication networks thus provide another high-performance, bidirectional transport environment to transparently convey digital video content. There are two possibilities of video content observation at the user-end. Observations can be carried out: on stationary or mobile computer monitor or, (in the case of using the home set-top-box, if needed) on the TV set screen. The picture quality delivered over IP networks depends on many factors, such as: source quality,

network parameters and network end-to-end bandwidth occupancy, a signal processing and performance of a display device.

The source quality depends on: method of coding, picture definition and signal bit rate, which are closely connected together. For preserving a good picture quality when picture definition is higher, the bit rates also have to be higher, for example:

Picture definition	Bit rate	Frame/ps
192x144	(24 to 52) kbit/s	12.5
192x144	141 kbit/s	12.5
384x288	273 kbit/s	12.5
384x288	(331 to 755) kbit/s	25
704x576	1119 kbit/s	25

The noise introduced in IP packet networks is described by the following parameters: packet loss ratio, latency, jitter (latency variation). When the good quality video content is introduced into the network according to the maximum end-to-end bandwidth and the packet rate available for video services, the network parameters required for preserving a good quality of the image should have values outlined and motivated in [ITU-R BT.1720]. The recommended objective measurement methods of the above parameters are also given in this Recommendation.

Each IP transport network should have defined maximum bandwidth per stream or packet rate for a given packet size, maximum packet rate per stream and maximum number of sustainable streams. The determination of the service quality, based on end-to-end measurements should provide the information on picture quality offered to the user. It depends to a high degree on network occupancy. If, in a network with a defined packet rate per stream, the number of sustainable streams by other services is too high, the service quality will be poor or not available.

This Recommendation deals with subjective picture quality assessment at the user end in home on standby or mobile computer monitor as well as on TV set screen. Conventional methods of laboratory testing do not replicate these conditions. It was considered useful, therefore, for the subjective quality of video delivered over IP networks to be measured continuously with subject viewing the material once without a source reference as in single stimulus method (SSCQE) but with some modifications resulted from specific service. This method is expected to be as good as the laboratory method, but has different goals. SSCQE was chosen as the basis of this Recommendation (e.g., rather than the absolute category rating (ACR) methodology), because SSCQE is a continuous quality method and can be used for longer test sequences. That is, in SSCQE as with home television viewing, video is presented continuously, without breaks between content.

5.2 Recommendations

5.2.1 Viewing conditions

Measurements will be made at different network's user-ends in home environment. Viewing conditions should be chosen to be similar to, but not more critical than, typical conditions in home environment given in [ITU-R BT.1129-2].

Condition	Item	Values
a	Ratio of viewing distance to picture height	6 (Note)
b	Peak luminance	200 cd/m²
c	Environmental Illuminance on the screen (Incident light from the environment falling on the screen should be measured perpendicularly on the screen)	200 Lux
NOTE – Other viewing environment variables are intentionally uncontrolled (e.g., wall colour). Reflections on the viewing screen should be avoided and the monitor should be set with the proper aspect ratio.		

5.2.2 Observers

In principle, 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 (see Note). Prior to a session, the observers should be screened for (corrected-to-) normal visual acuity and for normal colour vision using specially selected charts.

When estimations are made on computer screen, the number of simultaneous assessors may be 1 to 3 people at the same time, depending on the room and screen size. All observers should watch in the same conditions because reproduced colour and brighness may change according to the viewing position.

Computer screen resolution and brightness should be the same in all test presentation and should be given in test protocol.

Identical test sessions should be performed with all observers. That is, this identical session should be performed in the same place using the same test computer and on the same test material stored in computer memory.

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

5.2.3 Instruction for the observers

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.

In the case of picture evaluation (with accompanying audio), impression of video quality may be affected by audio quality, and observers should be instructed to evaluate the overall video quality.

5.2.4 Grading scale

The quality evaluation method uses a continuous quality scale to provide a measurement of the intrinsic quality of video sequences.

The digital television compression will produce impairments to the picture quality which are scene-dependent and time-varying. Even within short extracts of digitally-coded video, the quality can fluctuate quite widely depending on scene content, and impairments may be very short-lived.

Each observer moves a slider on a continuous scale graded from 0 to 100 annotated by 5 quality items linearly arranged (excellent, good, fair, poor, bad). During each program segment, every 0.5 second, samples are taken of from observers' sliders and sent to test computer with information about program content.

5.2.5 Test pictures

As evaluation concerns picture quality received in the home, 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 program hours. For example, half of items are definitely critical, and half of them are moderately critical.

Tests should be made on pictures of normal program hours as well as test pictures recommended by Table 1 of [ITU-R BT.1210-3].

5.2.6 The test session

Before the test, the trial session should be performed to evaluate that observers get the same viewing conditions (for example, using the pictures stored in the memory).

Before the right picture quality assessment is delivered to the home, a set of "practice presentations" should be introduced to stabilize the observers' opinion. It should last about 3 minutes and contains few program segments similar but not the same which will be shown later during the session.

The preliminary presentations are not to be taken into account in the statistical analysis of the test results. A right session, that is a series of presentations, should not last more than half an hour.

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

– *Program segment (PS)*: a PS corresponds to one program 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 min long;

– *Test session (TS)*: a TS is a series of one or more different combinations of 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 min duration;

– *Test presentation (TP)*: a TP represents the full performance of a test. A TP can be divided in TSs 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 is limited, a TP can be made of a repetition of the same TS to perform the test on a long enough period of time.

In the case of picture evaluation (with accompanying audio), selection of the accompanying audio material should be considered at the same level of importance as the selection of video material, prior to the test performance. Acoustic conditions should provide the same quality of audio to all observers.

The simplest test format would use a single PS and a single QP.

5.2.7 Presentation of the results

The results must be given together with the following information:

– observer data;

– data and time of observation;

– location of observation;

– objectively measured network's parameters;

– type of observed program segments;

– opinion (quality ratings) of several segments.

Data should be collated from all test sessions. 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 program segment, quality parameter or per entire test session.

Nevertheless, the varying delay in different viewer response time may influence the assessment results if only the average over a program segment is calculated. Analysis is carried out to evaluate the impact of the response time of different viewers on the resulting quality grade.

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

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.

Methods of analysis and presentation of results are given in Annex A.

Annex A

Methods of analysis and presentation of results

(This annex forms an integral part of this Recommendation)

A.1 Scope

This annex specifies the statistical techniques to yield results of subjective tests of picture quality delivered to home over broadband IP networks in numerical algorithm form. This algorithm concerns calculation of mean scores and calculation of confidence interval and summarizes the performance of the system under test.

This annex is intended, in particular, for network designers and operators. It gives them guidelines for network design taking into account all technical parameters as well as the number of sustainable streams and packet rate available for video services as well as the methodology of picture and service subjective measurements.

It may be also very helpful for users' complaints.

A.2 Introduction

In the course of a subjective test session to assess the performance of a television transmission system over broadband IP networks, a large amount of data is collected. These data, in the form of observers' electronic score sheets, must be condensed by statistical techniques to yield results in numerical and/or graphical algorithm (formulae) form which summarize the performance of the system under test.

In the method described in [ITU-R BT.500-11] continuous rating scale is used and the results are normalized to integer values between 0 and 100. There will be variations in the distribution of values due to differences in judgment between observers and the effect of the different programs observed.

A test will consist of a number of programs, L. Each program consists of a number of pictures, J which are divided into a number of presentations, corresponding to every 0.5 second, in which samples are taken from observers, K.

A.3 Calculation of mean scores

The first step is to analyse the results by calculating the mean score for each of the programs:

$$\bar{u}_{jk} = \frac{1}{N}\sum_{i=1}^{N} u_{ijk}$$

where:

u_{ijk} score of observer *i* for picture *j*, taken from presentation *k*

N number of observers

Similarly, overall mean scores, $\bar{u}_j$ and $\bar{u}_k$ could be calculated for each picture and each presentation, corresponding to every 0.5 second in which samples are taken from observers.

A.4 Calculation of confidence interval

The confidence interval is derived from the associated standard deviation. A standard deviation, S, could be calculated for each picture. It may be noted, however, that this standard deviation will, in cases where a small number of pictures is used, be influenced more by the differences between the observed picture than by the variations between the assessors participating in the assessment.

The standard deviation for each program is given by:

$$S_{jk} = \sqrt{\sum_{i=1}^{N} \frac{\left(\bar{u}_{jk} - u_{ijk}\right)^2}{(N-1)}}$$

It is proposed to use the 95% confidence interval which is given by:

$$\left[\bar{u}_{jk} - \delta_{jk,} \bar{u}_{jk} + \delta_{jk}\right]$$

$$\delta_{jk} = 1.96 \frac{S_{jk}}{\sqrt{N}}$$

With a probability of 95%, the absolute value of the difference between the experimental mean score and the "true" mean score (for a very high number of observers) is smaller than the 95% confidence interval, on condition that the distribution of the individual scores meets certain requirements.

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series D	General tariff principles
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
Series G	Transmission systems and media, digital systems and networks
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	**Cable networks and transmission of television, sound programme and other multimedia signals**
Series K	Protection against interference
Series L	Construction, installation and protection of cables and other elements of outside plant
Series M	Telecommunication management, including TMN and network maintenance
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	Telephone transmission quality, telephone installations, local line networks
Series Q	Switching and signalling
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks, open system communications and security
Series Y	Global information infrastructure, Internet protocol aspects and next-generation networks
Series Z	Languages and general software aspects for telecommunication systems

Printed in Switzerland
Geneva, 2009