

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

P.913

(06/2021)

SERIES P: TELEPHONE TRANSMISSION QUALITY,
TELEPHONE INSTALLATIONS, LOCAL LINE
NETWORKS

Audiovisual quality in multimedia services

**Methods for the subjective assessment of video
quality, audio quality and audiovisual quality of
Internet video and distribution quality television
in any environment**

Recommendation ITU-T P.913

ITU-T P-SERIES RECOMMENDATIONS

TELEPHONE TRANSMISSION QUALITY, TELEPHONE INSTALLATIONS, LOCAL LINE NETWORKS

Vocabulary and effects of transmission parameters on customer opinion of transmission quality	P.10–P.19
Voice terminal characteristics	P.30–P.39
Reference systems	P.40–P.49
Objective measuring apparatus	P.50–P.59
Objective electro-acoustical measurements	P.60–P.69
Measurements related to speech loudness	P.70–P.79
Methods for objective and subjective assessment of speech quality	P.80–P.89
Voice terminal characteristics	P.300–P.399
Objective measuring apparatus	P.500–P.599
Measurements related to speech loudness	P.700–P.709
Methods for objective and subjective assessment of speech and video quality	P.800–P.899
Audiovisual quality in multimedia services	P.900–P.999
Transmission performance and QoS aspects of IP end-points	P.1000–P.1099
Communications involving vehicles	P.1100–P.1199
Models and tools for quality assessment of streamed media	P.1200–P.1299
Telemeeting assessment	P.1300–P.1399
Statistical analysis, evaluation and reporting guidelines of quality measurements	P.1400–P.1499
Methods for objective and subjective assessment of quality of services other than speech and video	P.1500–P.1599

For further details, please refer to the list of ITU-T Recommendations.

Recommendation ITU-T P.913

Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment

Summary

Recommendation ITU-T P.913 describes non-interactive subjective assessment methods for evaluating the one-way overall video quality, audio quality and audiovisual quality for applications such as Internet video and distribution quality video. These methods can be used for several different purposes including, but not limited to, comparing the quality of multiple devices, comparing the performance of a device in multiple environments, and for subjective assessment where the quality impact of the device and the audiovisual material is confounded.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T P.913	2014-01-13	9	11.1002/1000/12106
2.0	ITU-T P.913	2016-03-15	9	11.1002/1000/12775
3.0	ITU-T P.913	2021-06-13	12	11.1002/1000/14704

Keywords

Audio quality, audiovisual quality, distribution quality video, Internet video, subjective assessment, video quality.

* To access the Recommendation, type the URL <http://handle.itu.int/> in the address field of your web browser, followed by the Recommendation's unique ID. For example, <http://handle.itu.int/11.1002/1000/11830-en>.

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/software copyrights, 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 appropriate ITU-T databases available via the ITU-T website at <http://www.itu.int/ITU-T/ipr/>.

© ITU 2021

All rights reserved. No part of this publication may be reproduced, by any means whatsoever, without the prior written permission of ITU.

Table of Contents

		Page
1	Scope.....	1
	1.1 Limitations.....	1
2	References.....	1
3	Definitions	2
	3.1 Terms defined elsewhere	2
	3.2 Terms defined in this Recommendation.....	2
4	Abbreviations and acronyms	3
5	Conventions	4
6	Source stimuli	4
	6.1 Source signal recordings.....	4
	6.2 Video considerations	5
	6.3 Audio considerations	7
	6.4 Audiovisual considerations	7
	6.5 Duration of stimuli	7
	6.6 Number of source stimuli	8
7	Test methods, rating scales and allowed changes.....	8
	7.1 List of methods	8
	7.2 Acceptable changes to the methods.....	11
	7.3 Discouraged but acceptable changes to the methods	12
8	Environment	12
	8.1 Controlled environment.....	13
	8.2 Public environment.....	13
	8.3 Viewing distance	13
9	Subjects.....	13
	9.1 Number of subjects.....	13
	9.2 Subject population	14
	9.3 Sampling subjects.....	14
	9.4 Sampling techniques.....	14
10	Experimental design	15
	10.1 Size of the experiment and subject fatigue.....	15
	10.2 Special considerations for transmission error, rebuffering and audiovisual synchronization impairments.....	15
	10.3 Special considerations for longer stalling events	16
	10.4 Pre-pilot testing and pilot testing.....	16
	10.5 Study design	16

	Page
11 Experiment implementation.....	17
11.1 Informed consent	17
11.2 Overview of subject screening	18
11.3 Optional pre-screening of subjects	18
11.4 Post-screening of subjects	19
11.5 Instructions and training	19
11.6 Study duration, sessions and breaks	20
11.7 Stimuli play mechanism	21
11.8 Voting	24
11.9 Questionnaire or interview	25
12 Data analysis.....	25
12.1 Documenting the experiment	25
12.2 Calculate MOS or DMOS	25
12.3 Evaluating objective metrics	26
12.4 Significance testing, subject bias and standard deviation of scores	26
12.5 Ratings from multiple laboratories	27
12.6 Improve MOS or DMOS data quality under challenging test conditions	27
13 Elements of subjective test reporting.....	29
13.1 Documenting the test design	29
13.2 Documenting the subjective testing.....	30
13.3 Data analysis.....	31
13.4 Additional information	31
Annex A – Method for post-experimental screening of subjects using Pearson linear correlation	32
A.1 Screen by PVS	32
A.2 Screen by PVS and HRC	33
Appendix I – Sample informed consent form	34
Appendix II – Sample instructions.....	35
Appendix III – Reference code for clause 12.6	36
Bibliography.....	42

Introduction

[ITU-T P.910], [b-ITU-T P.911] and [ITU-R BT.500-13] have been successfully used for many years to perform video quality and audiovisual quality subjective assessments. These Recommendations were initially designed around the paradigm of a fixed video service that transmits video over a reliable link to an immobile cathode ray tube (CRT) television located in a quiet and non-distracting environment, such as a living room or office. These Recommendations have been updated and expanded as technology shifted, and they have proved to be valuable and useful for the displays and questions addressed in their original scopes.

However, the initial premise of these Recommendations does not include the new paradigms of Internet video and distribution quality television. One new paradigm of video watching is an on-demand video service transmitted over an unreliable link to a variety of mobile and immobile devices located in a distracting environment, using liquid crystal displays (LCDs) and other flat-screen devices. This new paradigm impacts key characteristics of the subjective test, such as the viewing environment, the listening environment and the questions to be answered.

Users of Internet video and distribution quality television are moving from one device to another and from one environment to another throughout the day, perhaps even observing the same video using multiple devices. For example, someone might start watching a sporting event on their computer using Internet protocol television (IPTV), move to an over-the-air broadcast in their living room when the IPTV connection displays a rebuffering event and then switch to a mobile Internet device (MID) or even a smart phone when leaving the house. Thus, subjective quality assessments of Internet video and distribution quality television pose unique questions that are not considered in the existing Recommendations. These questions may require situation-specific modifications to the subjective scale (e.g., presentation of additional information defining what "good" means in this context).

Consider the pristine viewing environment defined by [ITU-R BT.500-13], with its exact lighting conditions and non-distracting walls. The intention is to remove the impact of the viewing and listening environment from the experiment. For some subjective audiovisual quality experiments, this is not appropriate. First, consider an experiment that investigates the quality of service observed by video-conferencing users in an office with fluorescent lights and the steady hum of a computer. Second, consider an experiment that analyses a communications device for emergency personnel. A highly distracting background may be a critical element of the experimental design (e.g., to simulate video watched inside a moving fire truck with sirens blaring). The impact of environment is an integral part of these experiments.

These questions and environments cannot be accommodated with the existing subjective assessment Recommendations. Modifying these Recommendations would reduce the value of the intended experiments and paradigms addressed therein. The main differences in this Recommendation when compared to existing ITU subjective assessment Recommendations are:

- 1) inclusion of multiple testing environment options (e.g., pristine laboratory environment, simulated office within a laboratory, public environment);
- 2) flexibility for the user to modify the subjective scale (e.g., modified words, added information);
- 3) applicability for interaction effects that confound the data (e.g., evaluating a device that can only accept compressed material, impact of mobility on quality perception);
- 4) mandatory reporting requirement (e.g., choices made where this Recommendation allows for flexibility, experimental variables that cannot be separated due to the experiment design); and
- 5) inclusion of multiple display technologies (e.g., flat screen, 2D, 3D).

Recommendation ITU-T P.913

Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment

1 Scope

This Recommendation describes methods to be used for subjective assessment of the audiovisual quality of Internet video and distribution quality. This may include assessment of visual quality only, audio quality only or the overall audiovisual quality. This Recommendation may be used to compare audiovisual device performance in multiple environments and to compare the quality impact of multiple audiovisual devices. It is appropriate for subjective assessment of devices where the quality impact of the device and the material is confounded. It is appropriate for a wide variety of display technologies, including flat screen, 2D, 3D, multi-view and autostereoscopic.

The devices and usage scenarios of interest herein are Internet video and distribution quality television. The focus is on the quality perceived by the end user.

1.1 Limitations

This Recommendation does not address the specialized needs of broadcasters and contribution quality television. This Recommendation is not intended to be used in the evaluation of audio-only stimuli alone, but rather audiovisual subjective assessments that may or may not include audio-only sessions. Caution should be taken when examining adaptive streaming impairments, due to the slow variations in quality within one stimulus over a long period of time.

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.340] Recommendation ITU-T J.340 (2010), *Reference algorithm for computing peak signal to noise ratio of a processed video sequence with compensation for constant spatial shifts, constant temporal shift, and constant luminance gain and offset.*
- [ITU-T P.78] Recommendation ITU-T P.78 (1996), *Subjective testing method for determination of loudness ratings in accordance with Recommendation P.76.*
- [ITU-T P.800] Recommendation ITU-T P.800 (1996), *Methods for subjective determination of transmission quality.*
- [ITU-T P.800.2] Recommendation ITU-T P.800.2 (2016), *Mean opinion score interpretation and reporting.*
- [ITU-T P.910] Recommendation ITU-T P.910 (2008), *Subjective video quality assessment methods for multimedia applications.*
- [ITU-T P.916] Recommendation ITU-T P.916 (2016), *Information and guidelines for assessing and minimizing visual discomfort and visual fatigue from 3D video.*

- [ITU-T P.1401] Recommendation ITU-T P.1401 (2020), *Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models*.
- [ITU-R BS.1534-3] Recommendation ITU-R BS.1534-1 (2015), *Method for the subjective assessment of intermediate quality level of coding systems*.
- [ITU-R BT.500-13] Recommendation ITU-R BT.500-13 (2012), *Methodology for the subjective assessment of the quality of television pictures*.
- [ITU-R BT.1788] Recommendation ITU-R BT.1788 (2007), *Methodology for the subjective assessment of video quality in multimedia applications*.

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

3.1.1 modality [b-ITU-T X.1244]: In general usage, this term refers to the forms, protocols, or conditions that surround formal communications. In the context of this Recommendation, it refers to the information encoding(s) containing information perceptible for a human being. Examples of modality include textual, graphical, audio, video or haptical data used in human-computer interfaces. Multimodal information can originate from, or be targeted to, multimodal-devices. Examples of human-computer interfaces include microphones for voice (sound) input, pens for haptic input, keyboards for textual input, mice for motion input, speakers for synthesized voice output, screens for graphic/text output, vibrating devices for haptic feedback, and Braille-writing devices for people with visual disabilities.

3.1.2 subjective assessment (picture) [b-ITU-T J.144]: The determination of the quality or impairment of programme-like pictures presented to a panel of human assessors in viewing sessions.

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

3.2.1 coding complexity: The ease or difficulty of maintaining perceptual quality of a video sequence as encoding bandwidth drops. Coding complexity plays a crucial role in determining the amount of video compression that is possible and, consequently, the level of impairment that is suffered when the scene is transmitted over a fixed-rate digital transmission service channel.

3.2.2 diegetic sound: Sound produced by objects appearing in the video or in the film's world, but off-screen.

3.2.3 dominant modality: The modality that carries the main information (i.e., audio or video).

3.2.4 double stimulus: A quality rating method where the subject is presented with two stimuli; the subject then rates both stimuli in the context of the joint presentation (e.g., a rating that compares the quality of one stimulus to the quality of the other).

3.2.5 hypothetical reference circuit (HRC): A fixed combination of a video encoder operating at a given bit rate, network condition and video decoder. The term HRC is preferred when vendor names should not be identified.

3.2.6 least distance of distinct vision: The closest distance at which someone with normal vision (20/20 vision) can comfortably look at something. This is sometimes known as "reference seeing distance".

3.2.7 non-diegetic sound: Sound produced by objects outside of the film's world, such as a narrator's voice-over.

- 3.2.8 processed:** The reference stimuli presented through a system under test.
- 3.2.9 processed video sequence (PVS):** The impaired version of a video sequence.
- 3.2.10 reference:** The original version of each source stimulus. This is the highest quality version available of the audio sample, video clip or audiovisual sequence.
- 3.2.11 reference seeing distance:** The closest distance at which someone with normal vision (20/20 vision) can comfortably look at something. This is sometimes called "least distance of distinct vision".
- 3.2.12 sequence:** A continuous sample of audio, video or audiovisual content.
- 3.2.13 single stimulus:** A quality rating method where the subject is presented with one stimulus and rates that stimulus in isolation (e.g., a viewer watches one video clip and then rates it).
- 3.2.14 source:** The content material associated with one particular audio sample, video clip or audiovisual sequence (e.g., a video sequence depicting a ship floating in a harbour).
- 3.2.15 spatial information:** The amount of detail in a video, e.g., from high contrast edges, fine detail and textures.
- 3.2.16 stimulus:** Audio sequence, video sequence or audiovisual sequence.
- 3.2.16 subject:** A person who evaluates stimuli by giving an opinion.
- 3.2.18 temporal forgiveness:** Impairments in video material which are to some extent forgiven if poor quality video is followed by good quality video.
- 3.2.19 temporal information:** The amount of temporal change in a video sequence.
- 3.2.20 terminal:** Device or group of devices used to play the stimuli during a subjective experiment (e.g., a laptop with earphones, or a Blu-ray player with a liquid crystal display (LCD) monitor and speakers).

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

2D	two Dimensional
3D	three Dimensional
2G	second Generation
3G	third Generation
ACR	Absolute Category Rating
CCR	Comparison Category Rating
CRT	Cathode Ray Tube
DCR	Degradation Category Rating
DMOS	Differential Mean Opinion Score
DSCS	Double Stimulus Comparison Scale
DSIS	Double Stimulus Impairment Scale
DV	Differential Viewer scores
HDTV	High Definition Television
HRC	Hypothetical Reference Circuit
IPTV	Internet Protocol Television

LCD	Liquid Crystal Display
LDDV	Least Distance of Distinct Vision
LPCC	Linear Pearson Correlation Coefficient
MID	Mobile Internet Device
MOS	Mean Opinion Score
MUSHRA	Multi-Stimuli with Hidden Reference and Anchor points
PC	Pair Comparison
PSNR	Peak Signal to Noise Ratio
PVS	Processed Video Sequence
RGB	Red–Green–Blue
RSD	Reference Seeing Distance
SAMVIQ	Subjective Assessment of Multimedia Video Quality
SOS	Standard deviation Of Scores
SQAM	Sound Quality Assessment Material
SQCIF	Sub-Quarter Common Intermediate Format
TV	Television
YUV	luminance (Y) – blue luminance (U) – red luminance (V)

5 Conventions

None.

6 Source stimuli

In order to evaluate quality in various circumstances, the content should cover a wide range of stimuli. The stimuli should be selected according to the goal of the test and recorded on a digital storage system. When the experimenter is interested in comparing results from different laboratories, it is necessary to use a common set of source stimuli to eliminate a further source of variation.

The selection of the test material should be motivated by the experimental question addressed in the study. For example, the content of the test stimuli should be representative of the full variety of programmes delivered by the service under study (sport, drama, film, speech, music, etc.).

6.1 Source signal recordings

The source signal provides the reference stimuli and the input for the system under test.

The quality of the reference stimuli should be as high as possible. As a guideline, the video signal should be recorded in uncompressed multimedia files using one of the following two formats: YUV (4:2:2 or 4:4:4 sampling) or RGB (24 or 32 bits). Usually the audio signal is taken from a high quality audio production. The audio CD quality is often the reference (16 bits, 44.1 kHz) such as the sound quality assessment material (SQAM) from the European Broadcasting Union (EBU), but if possible audio masters with a minimum of 16 bits and 48 kHz are preferred.

See clause 11.5.1 for more information on compressing reference video recordings.

6.2 Video considerations

The selection of the source video is nearly as important to the success or failure of a subjective test as the selection of hypothetical reference circuits (HRCs). The criteria described in this clause must be considered when selecting source videos. This clause contains guidelines for selecting the pool of source videos for an experiment.

6.2.1 Deviating from these criteria

The goal of scene selection is to represent the vast pool of all possible videos with a small handful of scenes. The scene selection criteria given below represent important considerations for success. This advice represents years of experience and lessons learned from subjective tests that failed.

Innovation requires inside the box thinking that contradicts traditional rules and constraints. Thus, some subjective tests will require deviation from this advice, resulting in totally new techniques for selecting videos. In such cases, it is critical to think about the implications of each decision. Explicitly choose and identify new scene selection criteria. Make trade-offs intelligently. Think about the impact of those trade-offs after the analysis and results reporting.

The reporting for all subjective tests must describe issues where the video scene selection deviated from or contradicted the advice given here.

6.2.2 Coding complexity

Ideally, the source videos will span the full range of coding difficulty of the target application. Video scenes with low coding complexity are easy to film and are thus readily available. Video scenes with high coding complexity are important, yet can be tricky to obtain. A well-designed experiment will contain videos with various coding complexities (high, medium and low). Coding complexity plays a crucial role in determining the amount of video compression that is possible and, consequently, the level of impairment that is suffered when the scene is transmitted over a fixed-rate digital transmission service channel.

Coding complexity is mainly impacted by two parameters: spatial information (detail) and temporal information (motion). Spatial information increases with the detail or sharpness visible within each frame, e.g., from high contrast edges, fine detail and textures. A video with low spatial information will have large areas with identical pixel values. Temporal information increases in proportion to individual pixels that change value from one frame or field to the next. Temporal information does not correspond to moving objects, but rather changing pixels. For example, a black rectangle sliding across a plain white background has high temporal information at the leading and trailing edge of the rectangle and no temporal information elsewhere. [ITU-R BT.1788] contains simple metrics to estimate spatial information and temporal information.

6.2.3 Subject matter

Ideally, the subject matter of each sequence will be typical of the target application. A set of scenes that contains limited variety of subject matter can lead to boredom and may not accurately reflect the target application. If the subject matter does not match the target application, this must be mentioned during the reporting.

6.2.4 Production quality and aesthetics

Ideally, the production quality of the video sequences will match the target application. For tests that focus on broadcast video applications, it is important that the reference videos have contribution quality and excellent aesthetics. For tests that focus on consumer video cameras, it may be important to choose video that contains typical camera impairments. These appear different from impairments simulated in software.

Scene quality will be impacted by physical characteristics of the camera, filming environment and initial recording. If a scene has poor technical quality, then it may be difficult for subjects to detect added impairments from the HRCs.

Scene quality will be impacted by aesthetics. The rating method and subject instructions can attempt to mitigate the impact of aesthetics. Nonetheless, poor aesthetics will impact quality ratings and thus the data analysis and conclusions despite all efforts to the contrary. Preferably, video content with poor aesthetics is avoided. If video content with poor aesthetics is used, this must be mentioned in the reporting and this confounding factor considered during the data analysis.

6.2.5 Post-production effects and scene cuts

Post-production effects and scene cuts can cause different portions of the encoded video sequence to have different quality levels. This can confuse subjects (e.g., make the subject unsure how to rate the video). Depending upon the purpose of the experiment, it may be advisable to avoid such video sequences. For example, an experiment that focuses on particular subroutines within a codec would avoid scene cuts; while an experiment that focuses on end-user perception of a particular broadcast service would typically include some content with rapid scene cuts.

6.2.6 Unusual properties

Valuable information is obtained from unique scenes with extraordinary features. Such stimuli can stimulate anomalous behaviour in the transmission chain.

The following scene traits can interact in unique ways with a codec or a person's perception. Ideally, the scene pool will include most of the following traits:

- action in a small portion of the total picture;
- animation, graphic overlays and scrolling text;
- blurred background, with an in-focus foreground;
- camera pans;
- camera still (locked down on a tripod);
- camera tilted;
- camera zoom;
- colourful scene;
- flashing lights or other extremely fast events;
- jiggling or bouncing picture (e.g., handheld camera);
- multiple objects moving in a random, unpredictable manner;
- night or dimly lit scene;
- ramped colour (e.g., sunset);
- repetitious or indistinguishable fine detail (e.g., gravel, grass, hair, rug, pinstripes);
- rotational movement (e.g., a carousel or merry-go-round seen from above);
- sharp black/white edges;
- small amounts of analogue noise (e.g., camera gain from dim lighting);
- very saturated colours;
- visually simple imagery (e.g., black birds flying across a blue sky);
- water, fire or smoke (for unusual shapes and shifting patterns).

6.2.7 Novelty and convenience sampling

It is possible to over-train on particular source sequences. For this reason, it is important to include new and novel source sequences in each new experiment.

To select videos from a small set of content that is easily available to the experimenter is a form of convenience sampling. A wide variety of videos is readily available. The practice of convenience sampling is justified, except for the most cutting edge video technologies.

6.3 Audio considerations

When testing the overall quality of audiovisual sequences, but not speech comprehension, the speech need not be in a language understood by all subjects.

All audio samples should be normalized for a constant volume level (e.g., normalize between clips, leaving volume variations within each clip alone). The audio source should preferably include a variety of audio characteristics (e.g., both male and female speakers, different musical instruments, different dynamic ranges). The dynamic range of an audio signal plays a crucial role in determining the impact of audio compression.

Post-production effects and scene cuts can cause different portions of the encoded audio sequence to have different quality levels. This can confuse subjects (e.g., make the subject unsure how to rate the video). Depending upon the purpose of the experiment, it may be advisable to avoid such audio sequences.

Items have to be chosen to be realistic types of audio excerpts as much as possible, keeping in mind that they must remain as critical as possible as well (this means that transparency is not often achieved by established encoders when encoding these audio sequences).

6.4 Audiovisual considerations

Specific care has to be taken when choosing source stimuli for audiovisual quality subjective assessments, since some degradation may have different impacts according to the relationship between audio and video. Aspects that should be considered are as follows:

- Diegetic or non-diegetic sounds: Diegetic sounds are produced by objects appearing in the video (e.g., a person visible on the screen is talking) or in the film's world, but off-screen (e.g., traffic noise, crowd noise). Non-diegetic sounds include voice-overs and background music.
- Dominant modality (audio or video). For example, in TV news sequences, the main information is carried by the audio modality, whereas a sport sequence would be more characterized by a video dominant modality.

Both aspects have been shown to have an impact on audiovisual quality, see [b-Lassalle]. For example, the perception of de-synchronization between image and sound is influenced by diegetic aspects.

6.5 Duration of stimuli

The methods in this Recommendation are intended for stimuli that range from 5 to 20 s in duration. Sequences of 8–10 s are highly recommended. For longer durations, it becomes difficult for viewers to take into account all of the quality variations and score properly in a global evaluation. The temporal forgiveness effects become important when the time duration of a stimulus is high, see [b-Hands].

Extra source content may be required at the beginning and end of each source stimulus. For example, when creating a 10 s processed stimulus, the source might have an extra 2 s of content before and after, to give a total of 14 s. The purpose of the extra content is to allow the audio and video coders to stabilize, and prevent the propagation of unrelated content into the processed stimuli (e.g., after the occurrence of digital transmission errors). The extra content should be discarded during editing. This technique is advised when analysing hardware coders or transmission errors.

In order to limit the duration of a test, stimulus durations of 10 s to 1 min are preferred. Test duration limitation also diminishes subjects' fatigue.

6.6 Number of source stimuli

The number and type of test scenes are critical for the interpretation of the results of the subjective assessment. So, four to six scenes are enough, if the variety of content is respected. The audiovisual content must have an interest in audio and video separately and conjointly.

The number of audio excerpts is very important in order to get enough data for the interpretation of the test results. A minimum of five audio items is required with respect to the range of content that can be encountered in "real life" (i.e., when using the systems under test).

The number of five items is also a good compromise in order to limit the duration of the test.

7 Test methods, rating scales and allowed changes

This clause describes the test methods, rating scales and allowable deviations. The method controls the stimuli presentation. The rating scale controls the way that people indicate their opinion of the stimuli. A list of appropriate changes to the method follows in this clause.

In-force and superseded versions of [ITU-T P.800], [ITU-T P.910] and [ITU-R BT.500-13] include alternate names for some test methods described in this clause. These alternate names are identified and may be used.

7.1 List of methods

This clause contains a listing of appropriate subjective test methods and rating scales.

7.1.1 Absolute category rating method

The absolute category rating (ACR) method is a category judgement where the test stimuli are presented one at a time and rated independently on a category scale. ACR is a single stimulus method. The subject observes one stimulus and then has time to rate that stimulus.

The ACR method uses the following five-level rating scale:

- 5 Excellent
- 4 Good
- 3 Fair
- 2 Poor
- 1 Bad

The numbers may optionally be displayed on the scale.

7.1.1.1 Comments

The ACR method produces a high number of ratings in a brief period of time.

ACR ratings confound the impact of the impairment with the influence of the content upon the subject (e.g., whether the subject likes or dislikes the production quality of the stimulus).

7.1.2 Degradation category rating or double stimulus impairment scale method

The degradation category rating (DCR) method presents stimuli in pairs. The first stimulus presented in each pair is always the reference. The second stimulus is that reference stimulus after processing by the systems under test. DCR is a double stimulus method. The DCR method is also known as the double stimulus impairment scale (DSIS) method.