

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

P.919

(10/2020)

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

Audiovisual quality in multimedia services

Subjective test methodologies for 360° video on head-mounted displays

Recommendation ITU-T P.919

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.919

Subjective test methodologies for 360° video on head-mounted displays

Summary

Recommendation ITU-T P.919 describes subjective assessment methods for evaluating quality of experience of short (between 10 s and 30 s) 360° videos. Recommendation ITU-T P.919 also outlines the characteristics of the source sequences to be used, such as duration, type of content and number of sequences. Details within Recommendation ITU-T P.919 are expected to change in subsequent editions, based on experiments into how best to conduct subjective tests with 360° content.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T P.919	2020-10-14	12	11.1002/1000/14429

Keywords

360° video, methodology, QoE, quality of experience, subjective test.

* 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/11.1002/1000/114429>.

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/>.

© ITU 2020

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
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	3
6	Selection of 360° source content.....	3
	6.1 Source signals recordings	4
	6.2 Spatial and temporal information	4
	6.3 Exploratory information	5
	6.4 Comfort and simulator sickness symptoms	5
	6.5 Duration of stimuli	5
	6.6 Audio considerations	5
7	Test methods	5
	7.1 Test methods for audiovisual quality	6
	7.2 Test methods for simulator sickness symptoms	7
	7.3 Test methods for exploration behaviour.....	8
	7.4 Methods to collect the observers' scores	8
8	Environment and equipment.....	9
	8.1 Test environment	9
	8.2 Equipment.....	9
9	Subjects.....	9
10	Experiment design	9
	10.1 Inclusion of reference conditions within the experiment	9
	10.2 Size of the experiment and subject fatigue	9
11	Experiment implementation.....	10
	11.1 Informed consent	10
	11.2 Viewer screening	10
	11.3 Post-screening of subjects	10
	11.4 Instructions and training	10
	11.5 Experiment sessions and breaks	11
	11.6 Questionnaire or interview	11
12	Data analysis.....	11
	12.1 Calculate mean opinion score or differential mean opinion score	12
	12.2 Analysis of exploration data	12
13	Elements of subjective test reporting.....	12

	Page
13.1 Documenting the test design	12
13.2 Documenting subjective testing	12
13.3 Data analysis.....	13
13.4 Additional information	13
Appendix I – Spatial and temporal information measurement for 360° video in the spherical domain.....	14
Appendix II – Computation of sample size from statistical power.....	16
Appendix III – Recommendations for information sheet and consent form.....	17
Appendix IV – Sample instructions	19
Appendix V – Sample questionnaire for background data on subjects	21
Appendix VI – Analysis of exploration data	22
Bibliography.....	29

Recommendation ITU-T P.919

Subjective test methodologies for 360° video on head-mounted displays

1 Scope

This Recommendation addresses the subjective evaluation of 360° video viewed with head-mounted displays (HMDs) that enable interaction of three degrees of freedom (3DoF) in head movement to explore content. As such, 360° video that may possibly include spatial audio presentation is different from traditional audiovisual media such as television or movies. 360° Videos are captured by cameras having a 360° field of view (FoV) and are thus able to capture the surrounding scene at each instant in time. Typically, when users view 360° videos on an HMD, they can turn around to view the immersive 360° space from different angles. Coupled with the large FoV presented by an HMD, 360° video can essentially provide a more immersive experience than that achievable with traditional video.

This Recommendation describes methods to evaluate aspects of quality of experience (QoE) for 360° video. In general, the test methods recommended in this Recommendation utilize a hierarchical design, where the entire evaluation process is divided into three abstraction layers. This Recommendation may be used to compare 360° viewing sessions where stimuli may differ in terms of, for example, the recording technique applied, the processing (such as projection schemes, coding or rendering-specific aspects), and the HMD devices used. This Recommendation describes subjective evaluation of short (between 10 s and 30 s) 360° videos. Topics include assessment methods, subjective scales, environmental conditions, equipment and data analysis. These experiments can assess phenomena such as audiovisual quality and simulator sickness.

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 P.800.2] Recommendation ITU-T P.800.2 (2013), *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.913] Recommendation ITU-T P.913 (2016), *Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment*.
- [ITU-R BT.500-14] Recommendation ITU-R BT.500-14 (2019), *Methodologies for the subjective assessment of the quality of television images*.
- [ITU-R BT.2420-1] Recommendation ITU-R BT.2420-1 (2020), *Collection of usage scenarios and current statuses of advanced immersive audio-visual systems*.

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

3.1.1 double stimulus [ITU-T P.913]: 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).

3.1.2 field of view (FoV) [b-ITU-R RS.1804]: The swath width and all areas covered when the instrument is scanned.

3.1.3 hypothetical reference circuit (HRC) [ITU-T P.913]: 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.

3.1.4 processed [ITU-T P.913]: The reference stimuli presented through a system under test.

3.1.5 processed video sequence (PVS) [ITU-T P.913]: The impaired version of a video sequence.

3.1.6 quality of experience (QoE) [b-ITU-T P.10]: The degree of delight or annoyance of the user of an application or service.

3.1.7 reference [ITU-T P.913]: The original version of each source stimulus. This is the highest quality version available of the audio sample, video clip or audiovisual sequence.

3.1.8 sequence [ITU-T P.913]: A continuous sample of audio, video or audiovisual content.

3.1.9 single stimulus [ITU-T P.913]: 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).

3.1.10 source [ITU-T P.913]: 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).

3.1.11 spatial information [ITU-T P.913]: The amount of detail in a video, e.g., from high contrast edges, fine detail and textures.

3.1.12 stimulus [ITU-T P.913]: Audio sequence, video sequence or audiovisual sequence.

3.1.13 subject [ITU-T P.913]: A person who evaluates stimuli by giving an opinion.

3.1.14 temporal information [ITU-T P.913]: The amount of temporal change in a video sequence.

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

3.2.1 head mounted display (HMD): A display worn on the body that fits over a user's head, which has small display optics in front of the eyes and is usually equipped with additional sensors to track the viewer's head motions such as coordinate positions, pitch, roll, and yaw. In some instances, the position of the user's gaze is also captured.

NOTE – Based on [ITU-R BT.2420-1],

3.2.2 refresh rate: The frequency with which a display updates an image.

3.2.3 simulator sickness: A physiological condition arising when exposed to a virtual reality environment

NOTE – Definition paraphrased from [b-Kennedy].

3.2.4 three degrees of freedom (3DoF): Programme material in which the user can freely look around in any direction (yaw, pitch, and roll). A typical use case is a user sitting in a chair looking at 3D VR/360° content on a head-mounted display.

NOTE – Based on [ITU-R BT.2420-1],

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

2D	two Dimensional
3D	three Dimensional
3DoF	3 Degrees of Freedom
ACR	Absolute Category Rating
ACR-HR	Absolute Category Rating with Hidden Reference
DCR	Degradation Category Rating
DMOS	Differential Mean Opinion Score
DSIS	Double Stimulus Impairment Scale
FoV	Field of View
HMD	Head-Mounted Device
MOS	Mean Opinion Score
PVS	Processed Video Sequence
QoE	Quality of Experience
RGB	Red–Green–Blue
SI	Spatial Information
SSQ	Simulator Sickness Questionnaire
TI	Temporal Information
VR	Virtual Reality
VRSQ	Virtual Reality Sickness Questionnaire
VSR	Vertigo Score Rating
YUV	luminance–blue luminance–red luminance

5 Conventions

None.

6 Selection of 360° source content

In order to evaluate 360° video quality and other terms defined in this Recommendation in various circumstances, the content should cover a wide range of stimuli. In particular, 360° content with a variety of spatial and temporal complexity, motion, and exploratory properties (in terms of focus of attention) should be used for accurate assessment.

The 360° videos 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. Examples of content types for 360° video experiments are provided in clause 4.2 of [ITU-R BT.2420-0].

This Recommendation covers the use of monoscopic 360° video content.

6.1 Source signals 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. If available, pristine, uncompressed reference files shall be used in the target maximum resolution both in terms of frame rate and spatial resolution, with 4:2:2 or 4:4:4 chroma encoding in the luminance (Y)–blue luminance (U)–red luminance (V) (YUV) colour space at a minimum of 8 bits, or in the red–green–blue (RGB) colour space at a total of 24 or 32 bits. The resolution of the source videos should be at least $3\,840 \times 1\,920$ pixels.

It is noted that for viewing 360° videos, the HMD for testing provides further constraints on the resolution of the source videos. To retain the immersive characteristics and guarantee a precise perception of the quality of panoramic videos, the subjective quality assessment should be conducted with HMDs rather than plane screen monitors. Immersion requires that the 360° content can fill the entire FoV in HMD. Unlike the display of traditional two dimensional (2D) video, which can be presented in a per-pixel manner on the screen with fixed size by adding black pixels to the low resolution video or showing only part of the high resolution content, the 360° video must be presented in its entirety despite different resolutions [b-Zhang].

6.2 Spatial and temporal information

The selection of test scenes is an important issue. In particular, the spatial and temporal perceptual information of the scenes are critical parameters. These parameters play 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. Relevant video test scenes must be chosen such that their spatial and temporal information is consistent with the video services that the digital transmission service channel is intended to provide. The set of test scenes should span the full range of spatial and temporal information of interest to users of the devices under test.

[ITU-T P.910] specifies simple metrics to estimate spatial information (SI) and temporal information (TI).

Existing measures for signal characterization are applied on planar representations of the 360° content (e.g., equirectangular or cube-map). The geometrical domain used to compute spatial and temporal indicators may have an influence in the characterization [b-DeSimone]. For example, the SI computed in equirectangular projection might take on a misleading perceptual characterization, due to the strong warping of the visual content around the poles. Similarly, considering the SI computed on the mosaicked cube-map planar images, vertical and horizontal edges corresponding to discontinuities at the borders between cube faces are taken into account, even if they are not features of the 360° content. To avoid this problem the computation of SI and TI can be done on each cube face separately considering the mean value across all faces as a measure of the spatial complexity of the entire 360° image.

Moreover, a single spatial complexity value might not be informative enough to characterize the entire 360° frames. Some content may show a significant variability in terms of SI and TI depending on the viewing direction (i.e., across cube faces), so the portion of 360° content attended by the user can have very different spatial complexity. To account for this variability, the variance of SI and TI over the cube faces can be considered. To select images having different characteristics, a suggested approach is to select images that have different variability across cube faces [b-DeSimone].

Finally, SI and TI can be calculated in the spherical domain. For each sampling point (m_i, n_i) on the equirectangular projected plane, it is first re-projected on to the sphere to get its corresponding longitude and latitude coordinates (φ_i, θ_i). The Sobel filter is then applied on the 3×3 window centred around the point on the sphere. Details of the computation of SI and TI in the spherical domain are provided in Appendix I.

How the spatial and temporal information is computed should be clearly reported.

6.3 Exploratory information

When viewing 360° videos, viewers explore the content (moving the head and eyes) according to the regions of interest in the presented videos. In the test, it should be ensured that the source sequences evoke a variety of different types of exploration behaviour.

The underlying property of videos is typically referred to as "saliency". With 360° content it can be obtained : a) from the positions of the head (and thus the centre of the FoV in the HMD) when it is referred to as "head saliency"; or b) from the positions of the eyes when using eye-trackers, when it is referred to as "head-eye saliency" [b-David]. In a test on audiovisual quality, source sequences that result in different saliency patterns should be used, from exploratory (nothing in the scene clearly catches the observers' attention) to focused contents (some objects stand out directing the observers' attention) [b-DeSimone].

For this purpose, subjective assessments by experts can be used, as well as objective measures for exploration or attention behaviours (e.g., similarity ring metric [b-ETSI TR 126 918] or entropy and inter-observer congruency [b-DeSimone]).

6.4 Comfort and simulator sickness symptoms

The content of the selected 360° sequences should be comfortable to watch and should not contain violent, sexual or disturbing content. Moreover, the selected sequences should not produce a high amount of simulator sickness in participants.

6.5 Duration of stimuli

The methods in this Recommendation are intended for stimuli that range from 10 s to 30 s in duration. Sequences of 10 s are recommended for assessing audiovisual quality.

Special attention should be paid when using sequences with time-varying properties (e.g., scene transitions). In addition, when dealing with non-uniform degradations (e.g., tile-based encoding), exploration patterns may change quality ratings between shorter and longer sequences.

6.6 Audio considerations

To evaluate video quality it is possible to use test stimuli either with or without audio. For video-only experiments, the missing audio should not be considered as an impairment.

When using non-uniform (e.g., tile-based encoding) degradations, audio coming from out of the region of interest (especially when using spatial audio) may influence audiovisual quality ratings.

7 Test methods

Measurement of the perceived quality of video requires the use of subjective scaling methods. The condition for such measurements to be meaningful is that there exists a relation between the physical characteristics of the stimulus, in this case the 360° video sequence presented to the subjects in a test, and the magnitude and nature of the sensation caused by the stimulus. The final choice of one of these methods for a particular application depends on several factors, such as the context, the purpose and where in the development process the test is to be performed.

Subjective experiments with 360° video may measure opinions on different perceptual scales:

- audiovisual quality;
- simulator sickness symptoms;
- exploration behaviour;
- presence, emotion response and other factors may be measured, but are not covered in this Recommendation.

These perceptual scales must be rated independently. This clause describes the test methods and rating scales. The method controls the sequence presentation. The rating scale controls the way that people indicate their opinion of the sequences.

7.1 Test methods for audiovisual quality

These methods are appropriate for evaluating video quality in subjective experiments on 360° video.

7.1.1 Absolute category rating

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

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 test method can be used when testing time is of relevance, since it produces a high number of ratings in a brief period of time [b-Singla].

7.1.2 Degradation category rating method

The degradation category rating (DCR) method presents sequences in pairs. The first stimulus presented in each pair is always the reference. The second stimulus is the same reference sequence after impairment 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. In this case, subjects are asked to rate the impairment of the second stimulus in relation to the reference. The following five-level scale for rating the impairment should be used:

- 5 imperceptible;
- 4 perceptible, but not annoying;
- 3 slightly annoying;
- 2 annoying;
- 1 very annoying.

The numbers may optionally be displayed on the scale.

7.1.2.1 Comments

The DCR method produces fewer ratings than ACR in the same period of time.

DSIS is statically more reliable than the ACR [b-Singla].