International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

G.1011

(07/2016)

SERIES G: TRANSMISSION SYSTEMS AND MEDIA, DIGITAL SYSTEMS AND NETWORKS

Multimedia Quality of Service and performance – Generic and user-related aspects

---

Reference guide to quality of experience assessment methodologies

Recommendation ITU-T G.1011

ITU-T

ITU-T G-SERIES RECOMMENDATIONS

TRANSMISSION SYSTEMS AND MEDIA, DIGITAL SYSTEMS AND NETWORKS

INTERNATIONAL TELEPHONE CONNECTIONS AND CIRCUITS	G.100–G.199
GENERAL CHARACTERISTICS COMMON TO ALL ANALOGUE CARRIER-TRANSMISSION SYSTEMS	G.200–G.299
INDIVIDUAL CHARACTERISTICS OF INTERNATIONAL CARRIER TELEPHONE SYSTEMS ON METALLIC LINES	G.300–G.399
GENERAL CHARACTERISTICS OF INTERNATIONAL CARRIER TELEPHONE SYSTEMS ON RADIO-RELAY OR SATELLITE LINKS AND INTERCONNECTION WITH METALLIC LINES	G.400–G.449
COORDINATION OF RADIOTELEPHONY AND LINE TELEPHONY	G.450–G.499
TRANSMISSION MEDIA AND OPTICAL SYSTEMS CHARACTERISTICS	G.600–G.699
DIGITAL TERMINAL EQUIPMENTS	G.700–G.799
DIGITAL NETWORKS	G.800–G.899
DIGITAL SECTIONS AND DIGITAL LINE SYSTEM	G.900–G.999
MULTIMEDIA QUALITY OF SERVICE AND PERFORMANCE – GENERIC AND USER-RELATED ASPECTS	**G.1000–G.1999**
TRANSMISSION MEDIA CHARACTERISTICS	G.6000–G.6999
DATA OVER TRANSPORT – GENERIC ASPECTS	G.7000–G.7999
PACKET OVER TRANSPORT ASPECTS	G.8000–G.8999
ACCESS NETWORKS	G.9000–G.9999

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

Recommendation ITU-T G.1011

Reference guide to quality of experience assessment methodologies

Summary

Recommendation ITU-T G.1011 provides a reference guide to existing standards for quality of experience (QoE) assessment methodologies. This Recommendation specifies QoE assessment approaches and classifications of different applications, and identifies a taxonomy of QoE assessment standards with different technical categorizations.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T G.1011	2010-06-29	12	11.1002/1000/10864
2.0	ITU-T G.1011	2013-05-14	12	11.1002/1000/11931
3.0	ITU-T G.1011	2015-06-29	12	11.1002/1000/12507
4.0	ITU-T G.1011	2016-07-29	12	11.1002/1000/12966

Keywords

QoE, Quality of experience.

* 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.

Table of Contents

Recommendation ITU-T G.1011

Reference guide to quality of experience assessment methodologies

1 Scope

Quality of experience (QoE) is a widely used term in the telecommunication industry. However, there is no single document available that describes methods for assessing QoE. This Recommendation provides a reference guide to existing standardized methods in ITU for QoE assessment. The overall approaches for the different assessment methodologies typically used are described, and a taxonomy of QoE assessment standards is defined.

For different applications, this Recommendation identifies suitable subjective assessment methodologies, and where applicable, objective assessment methods that can be used to estimate subjective opinion, in addition to giving guidance on their usage and limitations.

The intended users of this Recommendation include telecommunication carriers, service providers and equipment manufacturers.

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 G.107] Recommendation ITU-T G.107 (2015), *The E-model: a computational model for use in transmission planning*.

[ITU-T G.107.1] Recommendation ITU-T G.107.1 (2015), *Wideband E-model*.

[ITU-T G.1030] Recommendation ITU-T G.1030 (2014), *Estimating end-to-end performance in IP networks for data applications*.

[ITU-T G.1070] Recommendation ITU-T G.1070 (2012), *Opinion model for video-telephony applications*.

[ITU-T G.1071] Recommendation ITU-T G.1071 (2015), *Opinion model for network planning of video and audio streaming applications*.

[ITU-T J.140] Recommendation ITU-T J.140 (1998), *Subjective picture quality assessment for digital cable television systems*.

[ITU-T J.144] Recommendation ITU-T J.144 (2004), *Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference*.

[ITU-T J.246] Recommendation ITU-T J.246 (2008), *Perceptual visual quality measurement techniques for multimedia services over digital cable television networks in the presence of a reduced bandwidth reference*.

[ITU-T J.247] Recommendation ITU-T J.247 (2008), *Objective perceptual multimedia video quality measurement in the presence of a full reference*.

[ITU-T J.249] Recommendation ITU-T J.249 (2010), *Perceptual video quality measurement techniques for digital cable television in the presence of a reduced reference*.

[ITU-T J.341] Recommendation ITU-T J.341 (2016), *Objective perceptual multimedia video quality measurement of HDTV for digital cable television in the presence of a full reference*.

[ITU-T J.342] Recommendation ITU-T J.342 (2011), *Objective multimedia video quality measurement of HDTV for digital cable television in the presence of a reduced reference signal*.

[ITU-T J.343] Recommendation ITU-T J.343 (2014), *Hybrid perceptual bitstream models for objective video quality measurements*.

[ITU-T J.343.1] Recommendation ITU-T J.343.1 (2014), *Hybrid-NRe objective perceptual video quality measurement for HDTV and multimedia IP-based video services in the presence of encrypted bitstream data*.

[ITU-T J.343.2] Recommendation ITU-T J.343.2 (2014), *Hybrid-NR objective perceptual video quality measurement for HDTV and multimedia IP-based video services in the presence of non-encrypted bitstream data*.

[ITU-T J.343.3] Recommendation ITU-T J.343.3 (2014), *Hybrid-RRe objective perceptual video quality measurement for HDTV and multimedia IP-based video services in the presence of a reduced reference signal and encrypted bitstream data*.

[ITU-T J.343.4] Recommendation ITU-T J.343.4 (2014), *Hybrid-RR objective perceptual video quality measurement for HDTV and multimedia IP-based video services in the presence of a reduced reference signal and non-encrypted bitstream data*.

[ITU-T J.343.5] Recommendation ITU-T J.343.5 (2014), *Hybrid-FRe objective perceptual video quality measurement for HDTV and multimedia IP-based video services in the presence of a full reference signal and encrypted bitstream data*.

[ITU-T J.343.6] Recommendation ITU-T J.343.6 (2014), *Hybrid-FR objective perceptual video quality measurement for HDTV and multimedia IP-based video services in the presence of a full reference signal and non-encrypted bitstream data*.

[ITU-T P.561] Recommendation ITU-T P.561 (2002), *In-service non-intrusive measurement device – Voice service measurements*.

[ITU-T P.562] Recommendation ITU-T P.562 (2004), *Analysis and interpretation of INMD voice-service measurements*.

[ITU-T P.563] Recommendation ITU-T P.563 (2004), *Single-ended method for objective speech quality assessment in narrow-band telephony applications*.

[ITU-T P.564] Recommendation ITU-T P.564 (2007), *Conformance testing for voice over IP transmission quality assessment models*.

[ITU-T P.800] Recommendation ITU-T P.800 (1996), *Methods for subjective determination of transmission quality*.

[ITU-T P.800.1] Recommendation ITU-T P.800.1 (2016), *Mean opinion score (MOS) terminology*.

[ITU-T P.800.2] Recommendation ITU-T P.800.2 (2016), *Mean opinion score interpretation and reporting*.

[ITU-T P.805] Recommendation ITU-T P.805 (2007), *Subjective evaluation of conversational quality*.

[ITU-T P.806] Recommendation ITU-T P.806 (2014), *A subjective quality test methodology using multiple rating scales*.

[ITU-T P.830] Recommendation ITU-T P.830 (1996), *Subjective performance assessment of telephone-band and wideband digital codecs*.

[ITU-T P.835] Recommendation ITU-T P.835 (2003), *Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm*.

[ITU-T P.862] Recommendation ITU-T P.862 (2001), *Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs*.

[ITU-T P.862.1] Recommendation ITU-T P.862.1 (2003), *Mapping function for transforming P.862 raw result scores to MOS-LQO*.

[ITU-T P.862.2] Recommendation ITU-T P.862.2 (2007), *Wideband extension to Recommendation P.862 for the assessment of wideband telephone networks and speech codecs*.

[ITU-T P.863] Recommendation ITU-T P.863 (2014), *Perceptual objective listening quality assessment*.

[ITU-T P.910] Recommendation ITU-T P.910 (2008), *Subjective video quality assessment methods for multimedia applications*.

[ITU-T P.911] Recommendation ITU-T P.911 (1998), *Subjective audiovisual quality assessment methods for multimedia applications*.

[ITU-T P.920] Recommendation ITU-T P.920 (2000), *Interactive test methods for audiovisual communications*.

[ITU-T P.1201.1] Recommendation ITU-T P.1201.1 (2012), *Parametric non-intrusive assessment of audiovisual media streaming quality – Lower resolution application area*.

[ITU-T P.1201.2] Recommendation ITU-T P.1201.2 (2012), *Parametric non-intrusive assessment of audiovisual media streaming quality – Higher resolution application area*.

[ITU-T P.1202.1] Recommendation ITU-T P.1202.1 (2012), *Parametric non-intrusive bitstream assessment of video media streaming quality – Lower resolution application area*.

[ITU-T P.1202.2] Recommendation ITU-T P.1202.2 (2013), *Parametric non-intrusive bitstream assessment of video media streaming quality – Higher resolution application area*.

[ITU-T P.1301] Recommendation ITU-T P.1301 (2012), *Subjective quality evaluation of audio and audiovisual multiparty telemeetings*.

[ITU-T P.1302] Recommendation ITU-T P.1302 (2014), *Subjective method for simulated conversation tests addressing speech and audiovisual call quality*.

[ITU-T P.1311] Recommendation ITU-T P.1311 (2014), *Method for determining the intelligibility of multiple concurrent talkers*.

[ITU-T P.1312] Recommendation ITU-T P.1312 (2016), *Method for the measurement of the communication effectiveness of multiparty telemeetings using task performance*.

[ITU-T P.1501] Recommendation ITU-T P.1501 (2014), *Subjective testing methodology for web browsing*.

[ITU-R BS.1116-3]	Recommendation ITU-R BS.1116-3 (2015), *Methods for the subjective assessment of small impairments in audio systems*.
[ITU-R BS.1285]	Recommendation ITU-R BS.1285 (1997), *Pre-selection methods for the subjective assessment of small impairments in audio systems*.
[ITU-R BS.1387-1]	Recommendation ITU-R BS.1387-1 (2001), *Method for objective measurements of perceived audio quality*.
[ITU-R BS.1534-3]	Recommendation ITU-R BS.1534-3 (2015), *Method for the subjective assessment of intermediate quality level of audio systems*.
[ITU-R BT.500-13]	Recommendation ITU-R BT.500-13 (2012), *Methodology for the subjective assessment of the quality of television pictures*.

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following term defined elsewhere:

3.1.1 quality of experience [b-ITU-T P.10 Amd.2]: The overall acceptability of an application or service, as perceived subjectively by the end-user.

NOTE 1 – Quality of experience includes the complete end-to-end system effects (client, terminal, network, services infrastructure, etc.).

NOTE 2 – Overall acceptability may be influenced by user expectations and context.

3.2 Terms defined in this Recommendation

None.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

CIF	Common Intermediate Format
FR	Full Reference
HDTV	High Definition Television
HVGA	Half Video Graphics Array
IPTV	Internet Protocol Television
ISDN	Integrated Services Digital Network
MOS	Mean Opinion Score
$MOS_{A,V,MM}$	Mean Opinion Score for Audio, Video or Multimedia
NB	Narrowband
NR	No Reference
PSTN	Public Switched Telephone Network
QCIF	Quarter Common Intermediate Format
QoE	Quality of Experience
QVGA	Quarter Video Graphics Array
RR	Reduced Reference

SD	Standard Definition
SWB	Super Wideband
VGA	Video Graphics Array
VoIP	Voice over Internet Protocol
WB	Wideband

5 Conventions

This Recommendation uses the following conventions:

LAB laboratory testing;

MON monitoring;

PLN planning.

6 Overview of QoE assessment methods

This Recommendation gives a general overview of QoE assessment methods. For example, the concept of QoE assessment; different approaches to assessment, whether assessment is subjective or objective; target services, e.g., streaming video or web browsing; measurement approaches for different scenarios; and definitions of metrics for assessment models. Figure 6-1 shows the structure of the content in this Recommendation.

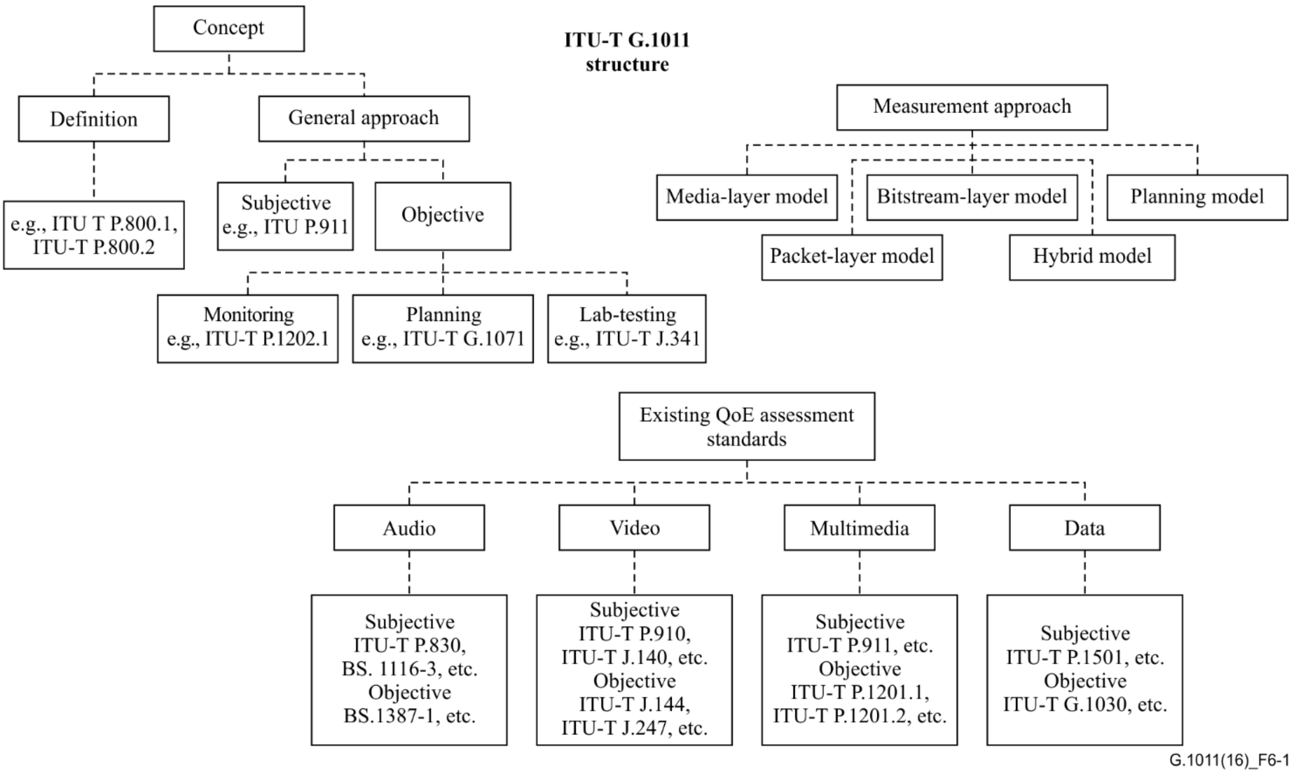

Figure 6-1 – Structure of the contents of this Recommendation

7 Concept

7.1 General approach to quality of experience assessment

In principle, assessment of quality of experience (QoE) must be performed using subjective tests with metrics, such as the mean opinion score (MOS). However, it is also possible and sometimes more convenient to estimate QoE based on objective testing and associated quality estimation models. Through different quality estimation models, it is possible to measure or calculate the objective parameters affecting QoE, in order to evaluate QoE. Subjective testing needs more resources and effort, because it requires human subjects. On the other hand, objective measurement and automatic calculation using appropriate quality estimation models is generally much faster and cheaper, but the accuracy of the final evaluation depends on the accuracy of these models.

In general, if QoE is evaluated objectively, it may be tested in intrusive mode, non-intrusive mode or planning mode. "Intrusive mode" means that the quality assessment system requires that a signal be injected into the system under test in order to generate a degraded output signal. This implies that the channel must be taken out of service for normal traffic. Conversely, for "non-intrusive mode", the quality assessment system can be used while live traffic is carried by the channel, without the need for any active test signals. "Planning mode" is not used in a real-time environment, but as a tool for the design of systems, and hence does not require any real-time inputs.

Irrespective of which mode it is in, the most important thing is to identify the key objective parameters that affect QoE and then relate these parameters to QoE measurement indices through specific quality estimation models.

Since QoE, i.e., perceived quality, is internal to the user, perceptual quality tests are ultimately the only means of validly and reliably assessing quality. Figure 7-1 gives a graphical explanation of the terms "valid" and "reliable".

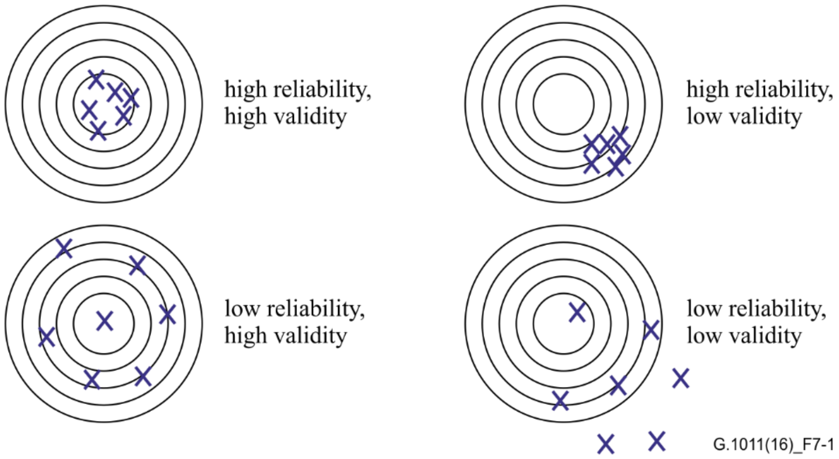

Figure 7-1 – Validity and reliability

NOTE – Validity describes how well a method actually measures what it is intended to measure. Reliability describes the accuracy of a method in terms of the scattering of results, e.g., when test assessment is repeated ([b-Guilford, 1954], pp. 349–357). Nowadays, validity is often replaced with the term accuracy, and reliability is replaced with the term precision.

The intrinsic variation that subjective test methods may be related with (e.g., from subject to subject or test laboratory to test laboratory) can be reduced based on a proper method design, yielding reliable and valid test results. There are a number of ITU Recommendations on how to properly design subjective tests, as summarized in clause 10.

Although objective assessment methods will lead to a certain level of reliability (i.e., the same quality estimate for a given input whenever it is applied to that input), their validity and reliability may be limited. Such model limitations are typically reflected in the corresponding Recommendations (see clause 10 for a summary of ITU Recommendations related to objective models). Within the range of system configurations, the objective methods that have been designed for, and proven to provide, valid and reliable quality predictions, can be used as a replacement for or a complement to subjective tests.

All methods can be classified according to how well diagnostic information can be obtained:

– Glass-box approaches enable insight into the internal functionality of the system. Such insight, in turn, may give access to diagnostic information. Hence, the source of a certain quality level may be identified and the relationship between system configuration and quality can be observed. For example, parameter-based models are typically glass-box models, since they build on a "diagnostic" description of the system.

– Black-box approaches do not enable insight into system internals. Hence, with black-box approaches direct indication of the source of low quality is not available. Certain signal-based models are typical examples. For black-box methods, the system under test is described only by its input and output. Based on additional processing of the results, diagnostic information may be inferred.

Current objective quality assessment techniques cannot necessarily be classified in a strict manner according to this black-box/glass-box dichotomy. Instead, different methods enable more or less diagnostic information to be obtained. Sometimes the terms "white box" and "grey box" are used in the literature in these cases ([b-Möller, 2005], pp. 86-89).

7.1.1 Subjective assessment

According to [b-Jekosch, 2005], p. 91, a (speech) quality test is defined as:

"[a] routine procedure for examining one or more empirically restrictive quality features of perceived speech with the aim of making a quantitative statement on these features".

Using this definition, subjective test methods can be broadly categorized as utilitarian and analytic test methods on the basis of whether the totality or a certain subset of perceived quality features is to be investigated (analytical), or whether single individual features or integral quality are to be measured (utilitarian).

Utilitarian test methods employ a unidimensional quality-rating scale, so that a direct comparison can be made between the qualities of different systems. In turn, analytical methods are related to a multidimensional analysis of test stimuli: They aim at identifying and/or quantifying some or all perceptual features underlying quality.

In order to capture the entire range of underlying auditory features, and to ensure a common understanding of these features, the reliability and validity of analytical methods can be significantly improved when trained listeners are used (e.g., [b-Mattila, 2001]; [b-Möller, 2000], pp. 105-120). In turn, tests of the utilitarian type are typically carried out with naive subjects. These should ideally represent the user group at which the system or service under investigation is aimed (age range, social and cultural background, etc.).

It should be noted that this utilitarian-analytical categorization of auditory test methods can be considered as being of utilitarian type itself, as it employs a unidimensional view on test methods. A multidimensional, that is, analytical view on speech quality assessment methods was provided by [b-Jekosch, 2005] (pp. 105-111).

Utilitarian methods have three main goals: "To be reasonably efficient in test administration and data analysis, to measure speech quality on a unidimensional scale, and to have good reliability in the test method" ([b-Quackenbush, 1988], pp. 15-16).

All test methods can be classified according to the categorization generally applicable to psychometric methods (after [b-Möller, 2000], pp. 48-49):

– The applied scaling method and scale level, with the scale levels being (after [b-Stevens, 1951], p. 25):

  1) Nominal (identifying equality);

  2) Ordinal (identifying greater or less);

3) Interval (identifying equality of intervals or differences);
4) Ratio (identifying equality of ratios).

– The presentation method used, such as:
  • Adjustment, i.e., the stimulus is changed to meet a certain description or judgement;
  • Constancy, i.e., one or several stimuli are presented per trial and a description or judgement is given (single stimulus, paired comparison, etc.).
– The test "modality":
  • Passive, e.g., listening or viewing;
  • Active, e.g., talking or moving;
  • Interactive, e.g., in a conversation.
– The instance at which quality is being judged:
  • Judgement after stimulus/stimuli presentation;
  • Continuous judgement during stimulus/stimuli presentation (e.g., in case of time-varying degradations).

7.1.2 Objective assessment

In order to reduce the necessity for time-consuming and costly perception tests to measure the quality of networks and systems, much effort has been spent on the development of alternative, instrumental, i.e., "objective" methods. It has to be noted, however, that the speech quality of today's telecommunication systems cannot readily be determined solely on the basis of basic signal measures like the signal-to-noise ratio. Although quality may be correlated with different instrumentally measurable signal characteristics, it is typically not possible to establish a simple relationship between these instrumentally measurable magnitudes and the quality perceived by a user of the system.

To account for the more complex interdependencies, different types of quality estimation and prediction models have been developed. Each model has its proper domain of application, and range of system or service conditions it has been designed for. Consequently, there is no universal quality model that can be applied in all circumstances. In order to better determine which model to choose for a given application, the models can be categorized according to different criteria (adapted from [b-Möller, 2002]):

– what the application is aimed at (network planning, codec optimization, service monitoring, etc.);
– the predicted quality features (integral quality – audio, video, audiovisual, speech, etc.), intelligibility, conversational aspects, etc.;
– the network components and configuration under consideration (entire connection end-to-end, codecs, etc.);
– the respective model input parameters (measured parameters like noise levels, information extracted from protocol headers like loss rates, entire signals, etc.);
– the way in which the input information is obtained:
  • by using passive measurements, which do not interfere with the running system or service, but involve an observation of the system or service during its normal operation;
  • by using active measurements, which involve dedicated measurement signals or data being processed by the system, and evaluating the transmitted result;
  • by estimation, e.g., during the network or service planning phase.
– the level of interactivity of the service aspects that can be assessed with the model, related to:
  • a passive perception (listening, viewing, etc.);

- an active usage (talking, moving);
- an interactive usage (e.g., for a conversation).

– the extent to which psychophysical knowledge or rather purely empirical data have been incorporated.

7.2 Content dependency of QoE assessment

QoE evaluation depends seriously on the context of subjective/objective experiments. [ITU-T P.800.1] provides various indices for MOS that clarify the context behind the MOS, which is one of the most widely used metrics. This metric is derived from the five-point ACR scale defined by several ITU-T P.800 series Recommendations (e.g., [ITU-T P.800]), ITU-T P.900 series Recommendations (e.g., [ITU-T P.910]) and [ITU-R BT.500-13]. For example, MOS-LQSW means that MOS was obtained in a "listening-only (one-way) test" (by letters "LQ"), and that the test was conducted in a subjective manner (by letter "S"), and that it was obtained for wideband (50-7000 Hz) speech relative to a wideband high quality reference.

[ITU-T P.800.2] gives additional details on how to provide contextual information regarding the reported MOS. For example, [ITU-T P.800.2] recommends reporting acoustic/viewing environment, audio/video material, and listening/viewing devices used in the experiment.

When one interprets the reported MOS, he or she needs to be careful about the context of an experiment, in which the MOS was derived.

8 Target services

This Recommendation gives guidelines for QoE assessment of various telecommunication services mainly utilizing audio and visual media.

8.1 Audio

– Conversational voice and voice messaging
  – Speech communication services, such as mobile telephony and voice over Internet protocol (VoIP), as well as conventional public switched telephone network (PSTN) services, are important targets of this Recommendation. The speech bandwidth can be either narrowband (NB) (300-3400 Hz) or wideband (WB) (100-7000 Hz).
– Streaming/on-demand audio
  – Audio delivery services provide audio content through telecommunication networks. They can be either download based or streaming based. Content is usually, but not limited to, musical material. Audio bandwidth can be from NB to full band (20-20000 Hz), although usually either super wideband (SWB) (50-14000 Hz) or full band (20-20000 Hz) is used.

8.2 Video

– Videophone/videoconferencing
  – Audiovisual communication refers to two-way interactive communication using speech and video signals, such as video telephony and videoconferencing. The bandwidth of speech signals can be either NB (300-3400 Hz) or WB (100-7000 Hz). Video resolution can range between quarter common intermediate format (QCIF) and high definition television (HDTV).
– Streaming/on-demand video
  – Video delivery services provide video content (normally with audio signals) through telecommunication networks. They can be either download based or streaming based. Typical content includes movies, documentaries, sports, news, and advertisement. Audio

bandwidth can be from NB (300-3400 Hz) to full band (20-20000 Hz), although usually either SWB (50-14000 Hz) or full band (20-20000 Hz) is used. Video resolution can be between QCIF and HDTV.

8.3 Data

– Data communication services targeted in this Recommendation are primarily web browsing.

9 Measurement approaches for different scenarios

Basically, there are five types of objective quality assessment methodologies. Depending on the application, these can be divided into media-layer models, packet-layer models, bitstream models, hybrid models, and planning models.

Objective quality assessment models can be used for several purposes:

– Planning: "Planning" refers to estimating the perceived quality of services of networks/systems before they are implemented. Since it is not used in a real-time environment, no real-time inputs are required to the objective model.
– Lab-testing: "Lab-testing" refers to estimating the perceived quality of services of networks/systems in the laboratory while the equipment is being developed.
– Monitoring: "Monitoring" refers to estimating the perceived quality of services of networks/systems that are operational. Necessary information is collected from the network and analysed to reflect the degradation of the quality experienced by users.

9.1 Media-layer models

Media-layer models (Figure 9-1) use the actual media signals (audio/video) as their input, and may take into account codec compression and channel characteristics. They use complex perceptually based psychophysical models to estimate QoE, by comparing (full reference/reduced reference (FR/RR)) the output (degraded) signal to the input (clean) signal or just analysing the output (degraded) signal (no reference (NR)).

The main applications of full reference models are QoE assessment in the laboratory, for example, codec comparison/optimization, since it uses not only the degraded (received) signal, but also the original (transmitted) signal to estimate QoE. Conversely, RR/NR models can be applied to QoE monitoring at the mid-point or end-points of the Internet protocol television (IPTV) network.

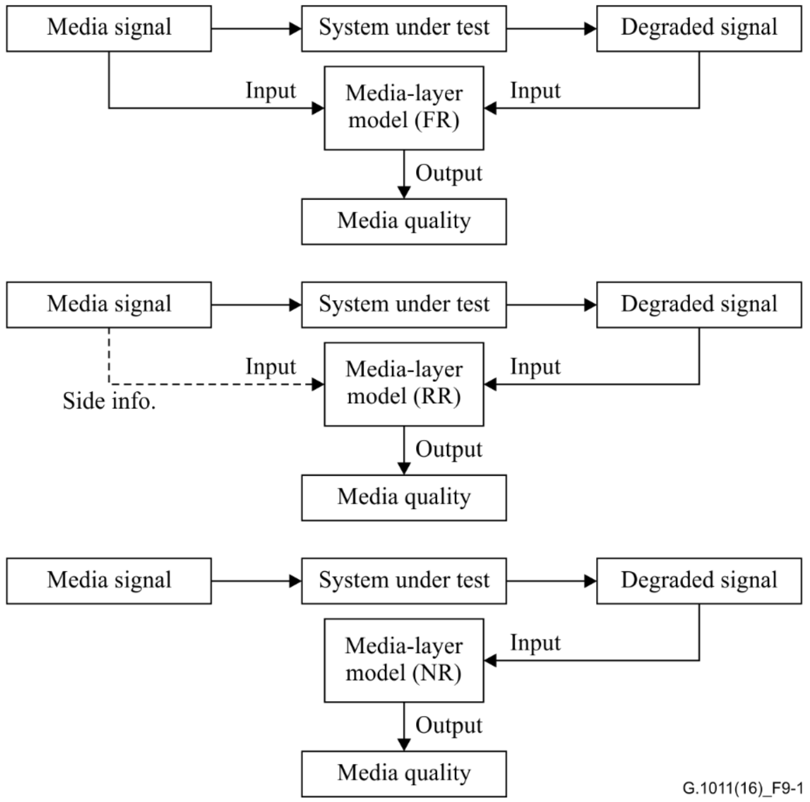

Figure 9-1 – Media-layer models

9.2 Packet-layer models

Packet-layer models (Figure 9-2) utilize only packet header information for QoE prediction. Because they do not parse the packet payload information, it is difficult to incorporate in such models aspects of QoE that are related to the media content, although they have a very light measurement of computational efficiency. These models are primary utilized as network probes at the mid-point or end-points of the network.

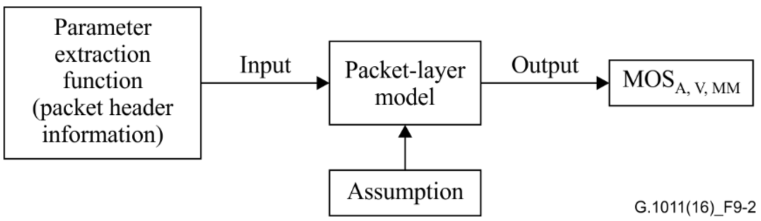

Figure 9-2 – Packet-layer model

9.3 Bitstream-layer models

A bitstream model (Figure 9-3) takes not only the encoded bitstream information, but also the packet header information as its input. Therefore, such a model can be viewed as a combination model between media-layer models and packet-layer models.

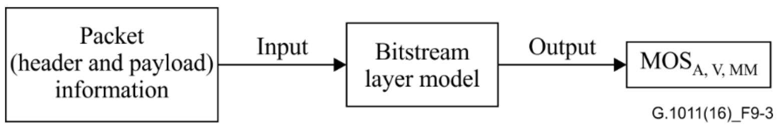

Figure 9-3 – Bitstream-layer model

Since the bitstream-layer model uses only the received packet information (degraded signal), it can be applied to QoE monitoring at the mid-point or end-points of the IPTV network.

9.4 Hybrid models

A hybrid model (Figure 9-4), as its name implies, is the combination of the previously mentioned models. It employs as much information as possible to predict QoE. An example of these models is shown below:

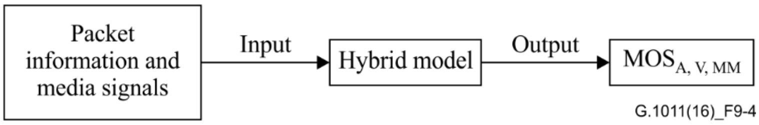

Figure 9-4 – Example of a hybrid model

9.5 Planning models

The input for planning models (Figure 9-5) includes the quality planning parameters of networks or terminals. It usually requires prior knowledge about the system under test. Such models can be applied to network planning and terminal/application design.

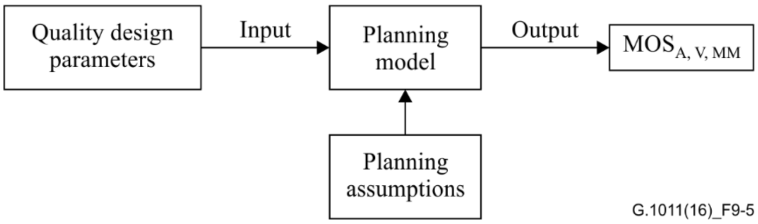

Figure 9-5 – Planning model

Standard examples of such models are [ITU-T G.107] for speech and [ITU-T G.1070] for videophone.

10 Taxonomy

Existing QoE assessment standards are categorized in Table 10-1. Appendix I provides information on planned standards.

Table 10-1 – Current ITU Recommendations for QoE assessment

Application	Media	Conversational (CONV)/Non-conversational (NONCONV)	Subjective test methodology	Objective test methodology			
				Model	**Transport protocol**	**FR/RR/NR**	**Primary usage**
Telephony	Speech	NONCONV	[ITU-T P.800] [ITU-T P.830] [ITU-T P.835] [ITU-T P.1301] [ITU-T P.1302] [ITU-T P.1311] [ITU-T P.806]	[ITU-T P.862] + [ITU-T P.862.1] (NB) [ITU-T P.862.2] (WB) [ITU-T P.863] (NB/WB/SWB)		FR	LAB, MON
				[ITU-T P.563] (NB) [ITU-T P.564] (NB/WB)		NR	MON
		CONV	[ITU-T P.800] [ITU-T P.805] [ITU-T P.1301] [ITU-T P.1312]	[ITU-T G.107] (NB) **[ITU-T G.107.1] (WB)**		NR	PLN
				[ITU-T P.561] + [ITU-T P.562] (NB/WB)		NR	MON
Video telephony	Multimedia (Note)	NONCONV	**[ITU-T P.1302**				
		CONV	[ITU-T P.920] [ITU-T P.1301]	[ITU-T G.1070] (NB/WB)		NR	PLN
Video streaming (Mobile TV/IPTV)	Video	NONCONV	[ITU-T P.910] [ITU-T J.140] [ITU-R BT.500-13]	[ITU-T J.144] standard definition (SD) [ITU-T J.247] (QCIF, common intermediate format (CIF), video graphics array (VGA)) [ITU-T J.341] (HD)	Transport agnostic	FR	LAB, MON

Table 10-1 – Current ITU Recommendations for QoE assessment

Application	Media	Conversational (CONV)/Non-conversational (NONCONV)	Subjective test methodology	Objective test methodology			
				Model	Transport protocol	FR/RR/NR	Primary usage
Video streaming (Mobile TV/IPTV)	Video	NONCONV	[ITU-T P.910] [ITU-T J.140] [ITU-R BT.500-13]	[ITU-T J.343.5] (HD, encrypted bitstream) [ITU-T J.343.6] (HD, non-encrypted bitstream)	UDP	FR	LAB, MON
				[ITU-T J.249] (SD) [ITU-T J.246] (QCIF, CIF, VGA) [ITU-T J.342] (HD)	Transport agnostic	RR	MON
				[ITU-T J.343.3] (HD, encrypted bitstream) [ITU-T J.343.4] (HD, non-encrypted bitstream)	UDP	RR	MON
				[ITU-T J.343.1] (HD, encrypted bitstream) [ITU-T J.343.2] (HD, non-encrypted bitstream)	UDP	NR	MON
	Audio	NONCONV	[ITU-T P.830] [ITU-R BS.1116-3] [ITU-R BS.1285] [ITU-R BS.1534-3]	[ITU-R BS.1387-1]	Transport agnostic	FR/RR	MON/PLN
	Multimedia	NONCONV	[ITU-T P.911]	[ITU-T P.1201.1] (QCIF, quarter video graphics array (QVGA), half video graphics array (HVGA)) [ITU-T P.1201.2] (SD, HD) [ITU-T P.1202.1] (QCIF, QVGA, HVGA) [ITU-T P.1202.2] (SD, HD)	UDP	NR	MON
				[ITU-T P.1201 Amd. 2], App. III	TCP	NR	MON
				[ITU-T G.1071] (SD, HD)	UDP	NR	PLN
Web browsing	Data		**[ITU-T P.1501]**	[ITU-T G.1030]	TCP	NR	PLN
	NOTE – For individual media (i.e., speech and video), the Recommendations used in telephony and video-streaming applications are applicable.						

Table10-2 gives detailed information on planning models, including existing and ongoing models.

Table 10-2 – Summary of existing planning models for video streaming applications

Transport protocol	Model	Video/Audiovisual	Video resolution	Video codec	Video type
UDP	ITU-T G.1071 (Lower resolution)	Video/Audiovisual	HVGA, QVGA, QCIF	MPEG-4 Part 2, H.264 (MPEG-4 Part 10)	
	ITU-T G.1071 (Higher resolution)	Video/Audiovisual	SD (PAL/NTSC), HD (720p50, 720p60, 1080i50, 1080p25,1080i60, 1080p30)	H.264 (MPEG-4 Part 10)	
TCP					

Table 10-3 gives detailed information on monitoring models, including existing and ongoing models.

Table 10-3 – Summary of existing monitoring models for video streaming applications

Transport protocol	Model	Video/Audiovisual	Video resolution	Video codec	Video type
UDP	ITU-T P.1201.1	Audiovisual	HVGA, QVGA, QCIF	MPEG-4 Part 2, H.264 (MPEG-4 Part 10)	IPTV
	ITU-T P.1201.2	Audiovisual	SD (PAL/NTSC), HD (720p50, 720p60, 1080i50, 1080p25,1080i60, 1080p30)	H.264 (MPEG-4 Part 10)	IPTV
	ITU-T J.343	Video	HD,VGA/WVGA	H.264/AVC (MPEG-4 Part 10)	IPTV
	ITU-T J.247	Video	QCIF，CIF，VGA	H.264/AVC (MPEG-4 part 10), VC-1, Windows Media 9, Real Video (RV 10), MPEG-4 Part 2	

Table 10-3 – Summary of existing monitoring models for video streaming applications

Transport protocol	Model	Video/Audiovisual	Video resolution	Video codec	Video type
	ITU-T J.246	Video	QCIF, CIF , VGA	H.264/AVC (MPEG-4 Part 10), VC-1, Windows Media 9, Real Video (RV 10), MPEG-4 Part 2.	
	ITU-T J.341	Video	1080i,1080p,1080i 50 Hz.1080p	H.264/AVC (MPEG-4 Part 10), MPEG-2	
TCP	ITU-T P.1201 Amd. 2, App. III	Audiovisual	HVGA, HD (1080i50, 1080p24, 1080i60, 1080p30)	H.264 (MPEG-4 Part 10)	Progressive downloading

Appendix I

(This appendix does not form an integral part of this Recommendation.)

Table I.1 provides information on existing and planned QoE assessment standards to give the users of this Recommendation an overall view of existing and ongoing projects in ITU. In addition, Table I.2 and Table I.3 provide indication of existing and planned QoE assessment standards for video streaming applications.

Table I.1 – Existing Recommendations and current work in ITU on QoE assessment

Application	Media	Conversational (CONV)/Non-conversational (NONCONV)	Subjective test methodology	Objective test methodology			
				Model	Transport protocol	FR/RR/NR	Primary usage
Telephony	Speech	NONCONV	[ITU-T P.800] [ITU-T P.806] [ITU-T P.830] [ITU-T P.835] [ITU-T P.1301] [ITU-T P.1302] [ITU-T P.1311]	[ITU-T P.862] + [ITU-T P.862.1] (NB) [ITU-T P.862.2] (WB) [ITU-T P.863] (NB/WB/SWB)		FR	LAB, MON
				[ITU-T P.563] (NB) [ITU-T P.564] (NB/WB)		NR	MON
		CONV	[ITU-T P.800] [ITU-T P.805] [ITU-T P.1301] [ITU-T P.1312]	[ITU-T G.107] (NB) [ITU-T G.107.1] (WB)		NR	PLN
				[ITU-T P.561] + [ITU-T P.562] (NB/WB)		NR	MON
Video telephony	Multimedia (Note)	NONCONV	[ITU-T P.1302				
		CONV	[ITU-T P.920] [ITU-T P.1301]	[ITU-T G.1070] (NB/WB)		NR	PLN

Table I.1 – Existing Recommendations and current work in ITU on QoE assessment

<table>
<tr><th rowspan="2">Application</th><th rowspan="2">Media</th><th rowspan="2">Conversational (CONV)/Non-conversational (NONCONV)</th><th rowspan="2">Subjective test methodology</th><th colspan="4">Objective test methodology</th></tr>
<tr><th>Model</th><th>Transport protocol</th><th>FR/RR/ NR</th><th>Primary usage</th></tr>
<tr><td rowspan="11">Video streaming (Mobile TV/IPTV)</td><td rowspan="5">Video</td><td rowspan="5">NONanCONV</td><td rowspan="5">[ITU-T P.910]
[ITU-T J.140]
[ITU-R BT.500-13]</td><td>[ITU-T J.144] standard definition (SD)
[ITU-T J.247] (QCIF, common intermediate format (CIF), video graphics array (VGA))
[ITU-T J.341] (HD)</td><td>Transport agnostic</td><td>FR</td><td>LAB, MON</td></tr>
<tr><td>[ITU-T J.343.5] (HD, encrypted bitstream)
[ITU-T J.343.6] (HD, non-encrypted bitstream)</td><td>UDP</td><td>FR</td><td>LAB, MON</td></tr>
<tr><td>[ITU-T J.249] (SD)
[ITU-T J.246] (QCIF, CIF, VGA)
[ITU-T J.342] (HD)</td><td>Transport agnostic</td><td>RR</td><td>MON</td></tr>
<tr><td>[ITU-T J.343.3] (HD, encrypted bitstream)
[ITU-T J.343.4] (HD, non-encrypted bitstream)</td><td>UDP</td><td>RR</td><td>MON</td></tr>
<tr><td>[ITU-T J.343.1] (HD, encrypted bitstream)
[ITU-T J.343.2] (HD, non-encrypted bitstream)</td><td>UDP</td><td>NR</td><td>MON</td></tr>
<tr><td>Audio</td><td>NONCONV</td><td>[ITU-T P.830]
[ITU-R BS.1116-3]
[ITU-R BS.1285]
[ITU-R BS.1534-3]</td><td>[ITU-R BS.1387-1]</td><td></td><td>FR/RR</td><td>MON/PLN</td></tr>
<tr><td rowspan="3">Multimedia</td><td rowspan="3">NONCONV</td><td rowspan="3">[ITU-T P.911]</td><td>[ITU-T P.1201.1] (QCIF, quarter video graphics array (QVGA), half video graphics array (HVGA))
[ITU-T P.1201.2] (SD, HD)
[ITU-T P.1202.1] (QCIF, QVGA, HVGA)
[ITU-T P.1202.2] (SD, HD)</td><td>UDP</td><td>NR</td><td>MON</td></tr>
<tr><td>[ITU-T P.1201 Amd. 2], App. III (HD)
P.NATS phase1(SD,HD)
P.NATS phase2 (SD,HD,WQHD,UHD)</td><td>TCP</td><td>NR</td><td>MON</td></tr>
<tr><td>[ITU-T G.1071] (SD, HD)
G.OM_HEVC (SD,HD)</td><td>UDP</td><td>NR</td><td>PLN</td></tr>
<tr><td>Web browsing</td><td>Data</td><td></td><td>[ITU-T P.1501]</td><td>[ITU-T G.1030]</td><td>TCP</td><td>NR</td><td>PLN</td></tr>
<tr><td colspan="8">NOTE – For individual media (i.e., speech and video), the Recommendations used in telephony and video-streaming applications are applicable.</td></tr>
</table>

Table I.2 – Summary of existing and ongoing planning models for video streaming applications

Transport protocol	Model	Video/Audiovisual	Video resolution	Video codec	Video type
UDP	ITU-T G.1071 (Lower resolution)	Video/Audiovisual	HVGA, QVGA, QCIF	MPEG-4 Part 2, H.264 (MPEG-4 Part 10)	
	ITU-T G.1071 (Higher resolution)	Video/Audiovisual	SD (PAL/NTSC), HD (720p50, 720p60, 1080i50, 1080p25,1080i60, 1080p30)	H.264 (MPEG-4 Part 10)	
	G.OM_HEVC	Video/Audiovisual	Phase1: HD	Phase1: ITU-T H.265 (MPEG-H Part 2)	IPTV
TCP					

Table I.3 – Summary of existing and ongoing monitoring models for video streaming applications

Transport protocol	Model	Video/Audiovisual	Video resolution	Video codec	Video type
UDP	ITU-T P.1201.1	Audiovisual	HVGA, QVGA, QCIF	MPEG-4 Part 2, H.264 (MPEG-4 Part 10)	IPTV
	ITU-T P.1201.2	Audiovisual	SD (PAL/NTSC), HD (720p50, 720p60, 1080i50, 1080p25,1080i60, 1080p30)	H.264 (MPEG-4 Part 10)	IPTV
	ITU-T J.343	Video	HD,VGA/WVGA	H.264/AVC (MPEG-4 Part 10)	IPTV
	ITU-T J.247	Video	QCIF, CIF, VGA	H.264/AVC (MPEG-4 part 10), VC-1, Windows Media 9, Real Video (RV 10), MPEG-4 Part 2	
	ITU-T J.246	Video	QCIF, CIF , VGA	H.264/AVC (MPEG-4 Part 10), VC-1, Windows Media 9, Real Video (RV 10), MPEG-4 Part 2.	
	ITU-T J.341	Video	1080i,1080p,1080i 50 Hz.1080p	H.264/AVC (MPEG-4 Part 10),MPEG-2	
TCP	ITU-T P.1201 Amd. 2, App. III	Audiovisual	HVGA, HD (1080i50, 1080p24, 1080i60, 1080p30)	ITU-T H.264 (MPEG-4 Part 10)	Progressive downloading
	ITU-T P.NATS phase1	Audiovisual	240p, 360p, 480p, 720p, 1080p	ITU-T H.264 (MPEG-4 Part 10) baseline, main, high, and high10 profiles	Progressive and adaptive downloading
	P.NATS phase2	Audiovisual	240p, 360p, 480p, 720p, 1080p,2K,4K	ITU-T H.264 (MPEG-4 Part 10) baseline, main, high, and high10 profiles H.265 baseline, main, high VP9	Progressive and adaptive downloading

Appendix II

(This appendix does not form an integral part of this Recommendation.)

In this appendix, common metrics are defined for QoE assessment models for video streaming applications, which could be transmitted via TCP or UDP.

II.1 Metrics related to TCP based streaming video

The following table defines common metrics for buffering in TCP based streaming video. These metrics could be used in OTT streaming video transmitted via HTTP/TCP, HLS/HTTP/TCP, DASH/HTTP/TCP.

Table II.1 – Definition for buffering in TCP based streaming video

Metrics	Definition
Initial buffering	Reference to definition in G.102y
Re-buffering	Reference to definition in G.102y
Stalling	Reference to definition in G.102y

II.2 Metrics related to UDP based streaming video

The following table defines common metrics for buffering in UDP based streaming video. These metrics could be used in streaming video transmitted via UDP, such as IPTV.

Table II.2 – Definition for buffering in UDP based streaming video

Metrics	Definition
Initial buffering	Reference to definition in [b-ITU-T G.1021]
Re-buffering	Reference to definition in [b-ITU-T G.1021]

Bibliography

[b-ITU-T G.1021] Recommendation ITU-T G.1021 (2012), *Buffer models for development of client performance metrics*.

[b-ITU-T P.10 Amd.2] Recommendation ITU-T P.10/G.100 (2006) Amendment 2 (2008), *New definitions for inclusion in Recommendation ITU-T P.10/G.100.*

[b-DSL Forum] DSL Forum TR-126 (2006), *Triple-play Services Quality of Experience (QoE) Requirements*.

[b-Guilford, 1954] Guilford, J.P. (1954), *Psychometric Methods*, 2nd edition. New York, NY: McGraw Hill. 597 p.

[b-Jekosch, 2005] Jekosch, U. (2005), *Voice and speech quality perception – Assessment and evaluation.* Berlin: Springer. 208 p.

[b-Mattila, 2001] Mattila, V.V. (2001), *Descriptive analysis of speech quality in mobile communications,* Vol. 340. Doctoral dissertation thesis, Tampere University of Technology, Tampere. See also AES E-Library, paper No. 5455 (viewed 2016-09-30) <http://www.aes.org/e-lib/browse.cfm?elib=9880>

[b-Möller, 2000] Möller, S. (2000), *Assessment and prediction of speech quality in telecommunications,* Boston, MA: Kluwer. 244 p.

[b-Möller, 2002] Möller, S., Raake, A. (2002). Telephone speech quality prediction: Towards network planning and monitoring models. *Speech Communication* Vol. 38: pp. 47-75 <http://www.sciencedirect.com/science/article/pii/S0167639301000437>

[b-Möller, 2005] Möller, S. (2005), *Quality of telephone-based spoken dialogue systems.* New York, NY: Springer. 469 p.

[b-Quackenbush, 1988] Quackenbush, S.R., Barnwell III, T.P., Clements, M.A. (1988). *Objective measures of speech quality.* Englewood Cliffs, NJ: Prentice Hall. 377 p.

[b-Stevens, 1951] Stevens, S.S. (1951). *Stevens' handbook of experimental psychology.* New York, NY: Wiley. 1436 p.

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	Environment and ICTs, climate change, e-waste, energy efficiency; 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	Terminals and subjective and objective assessment methods
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, Internet of Things and smart cities
Series Z	Languages and general software aspects for telecommunication systems

Printed in Switzerland
Geneva, 2016