International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

G.1070

(04/2007)

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

Quality of service and performance – Generic and user-related aspects

---

Opinion model for video-telephony applications

ITU-T Recommendation G.1070

ITU-T

International Telecommunication Union

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
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.

ITU-T Recommendation G.1070

Opinion model for video-telephony applications

Summary

ITU-T Recommendation G.1070 proposes an algorithm that estimates videophone quality to the QoE/QoS planners. This model can be used by QoE/QoS planners to help ensure that users will be satisfied with end-to-end service quality. Actually they want to avoid under-engineering. The outputs from the model are multimedia quality considering the interactive quality.

The application of this Recommendation is limited to QoE/QoS planning. Other applications such as quality benchmarking and monitoring are outside the scope of this Recommendation.

Source

ITU-T Recommendation G.1070 was approved on 22 April 2007 by ITU-T Study Group 12 (2005-2008) under the ITU-T Recommendation A.8 procedure.

Keywords

Multimedia, MOS, opinion model, QoE, QoS, quality planning, subjective quality, video telephony.

FOREWORD

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

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

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

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

NOTE

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

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

INTELLECTUAL PROPERTY RIGHTS

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

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

CONTENTS

ITU-T Recommendation G.1070

Opinion model for video-telephony applications

1 Scope

This Recommendation describes a computational model for point-to-point interactive videophone applications over IP networks that is useful as a QoE/QoS planning tool for assessing the combined effects of variations in several video and speech parameters that affect the quality of experience (QoE). This model can be used by QoE/QoS planners to help ensure that users will be satisfied with end-to-end service quality. Actually they want to avoid under-engineering. Network, application, and terminal quality parameters of high importance to QoE/QoS planners are incorporated into this model.

The model provided in this Recommendation needs to be a flexible tool capable of providing feedback on individual qualities as well as overall quality.

This Recommendation is very different from [ITU-T J.148] in terms of input parameters. In this Recommendation, multimedia quality is calculated by using network, application, and terminal equipment parameters, whereas in [ITU-T J.148], the calculation is done by using speech and video signals.

This Recommendation assumes videophone applications using dedicated videophone terminals, desktop PCs, laptop PCs, PDAs, and mobile phones. The speech bandwidth is limited to the telephone band (300-3400 Hz).

The application of this Recommendation is limited to QoE/QoS planning. Other applications such as quality benchmarking and monitoring are outside the scope of this Recommendation.

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] ITU-T Recommendation G.107 (2005), *The E-model, a computational model for use in transmission planning*.

[ITU-T G.113] ITU-T Recommendation G.113 (2001), *Transmission impairments due to speech processing*.

[ITU-T G.122] ITU-T Recommendation G.122 (1993), *Influence of national systems on stability and talker echo in international connections*.

[ITU-T G.711] ITU-T Recommendation G.711 (1988), *Pulse code modulation (PCM) of voice frequencies*.

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

[ITU-T J.148] ITU-T Recommendation J.148 (2003), *Requirements for an objective perceptual multimedia quality model*.

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

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

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

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

[ITU-T P.833] ITU-T Recommendation P.833 (2001), *Methodology for derivation of equipment impairment factors from subjective listening-only tests*.

[ITU-T P.834] ITU-T Recommendation P.834 (2002), *Methodology for the derivation of equipment impairment factors from instrumental models*.

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

[ITU-T P.862.2] ITU-T Recommendation P.862.2 (2005), *Wideband extension to Recommendation P.862 for the assessment of wideband telephone networks and speech codecs*.

[ITU-T P.910] ITU-T Recommendation P.910 (1999), *Subjective video quality assessment methods for multimedia applications*.

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

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

[ITU-R BS.1387-1] ITU-R Recommendation BS.1387-1 (2001), *Method for objective measurements of perceived audio quality*.

3 Definitions

This Recommendation defines the following terms as shown in Table 1:

Table 1 – List of definitions

Name	Description	Unit
AD	Absolute audiovisual delay	–
b_n	Video bit rate (n = 1, 2, …, N)	kbit/s
Bpl_S	Speech packet-loss robustness	–
Br_V	Video bit rate	kbit/s
D_{bnfm}	Degree of video quality robustness against packet loss (n = 1, 2, …, N, m = 1, 2, …, M)	–
D_{FrV}	Degree of video quality robustness due to frame rate reduction	–
D_n	Degree of video quality robustness due to frame rate reduction (n = 1, 2, …, N)	–

Table 1 – List of definitions

Name	Description	Unit
D_{PplV}	Degree of video quality robustness against packet loss	–
f_m	Frame rate (m = 1, 2, …, M)	fps
Fr_V	Video frame rate	fps
I_{coding}	Objective measurement of basic video quality accounting for coding distortion	–
Idd	Degradation caused by pure delay in ITU-T Rec. G.107	–
Idte	Degradation caused by talker echo	–
Ie-eff	Degradation caused by speech coding and packet loss	–
Ie_S	Speech coding distortion	–
I_n	Objective measurement of maximum video quality at each bit rate (n = 1, 2, …, N)	–
I_{Ofr}	Objective measurement of maximum video quality at each bit rate	–
MM_q	Objective measurement of multimedia quality accounting for influence of speech and video quality and speech and video delay	–
MM_{SV}	Audiovisual quality	–
MM_T	Audiovisual delay impairment factor	–
MS	Audiovisual media synchronization	–
Ofr	Optimal frame rate that maximizes video quality at each bit rate	–
O_n	Optimal frame rate (n = 1, 2, …, N)	–
Ppl_S	Speech packet-loss rate	%
Ppl_V	Video packet-loss rate	%
Q	Speech quality index	–
R	Transmission rating factor	–
S_q	Objective measurement of speech quality	–
$S_q(V_q)$	Objective measurement of speech quality accounting for influence of video quality	–
TELR	Talker echo loudness rating	dB
T_S	Speech delay	ms
T_V	Video delay	ms
V_q	Objective measurement of video quality	–
$V_q(S_q)$	Objective measurement of video quality accounting for influence of speech quality	–
V_{qs}	Subjective video quality	
$V_{qs}(b_n, f_m)$	Subjective video quality under conditions of b_n and f_m	–

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

ACR Absolute Category Rating

AEC Acoustic Echo Canceller

CIF Common Intermediate Format (352×288 pixels)

CT Codec Type

ERL Echo Return Loss

KFI Key Frame Interval

MOS Mean Opinion Score

QCIF Quarter CIF (176×144 pixels)

QQVGA Quarter Quarter VGA (160×120 pixels)

QVGA Quarter Video Graphics Array (320×240 pixels)

RLR Receiving Loudness Rating

SLR Sending Loudness Rating

VDS Video Display Size

VF Video Format

VGA Video Graphics Array (640×480 pixels)

5 Conventions

In this Recommendation, "subjective quality" refers to the MOS obtained in ACR tests as defined in [ITU-T P.800], [ITU-T P.910], [ITU-T P.911], and [ITU-T P.920], depending on the media under evaluation and evaluation context such as one-way listening/viewing and two-way interactive communication.

6 Application

The application of this Recommendation is limited to QoE/QoS planning. Other applications such as quality benchmarking and monitoring are outside the scope of this Recommendation. Table 2 gives an overview of various Recommendations related to objective quality assessment methods and their intended applications, the media types they apply to, and the subjective quality aspects that are taken into account.

Table 2 – Relationship of G.1070 to other ITU Recommendations for objective quality assessment

<table>
<tr><td rowspan="2">Media</td><td rowspan="2">Estimated subjective quality</td><td colspan="3">Application</td></tr>
<tr><td>Benchmarking/
Intrusive monitoring</td><td>Non-intrusive monitoring</td><td>Network planning</td></tr>
<tr><td rowspan="3">Speech</td><td>One way
(Listening quality)</td><td>P.862/P.862.1
(Telephone band)
P.862.2 (Wideband)</td><td>P.563, P.564
(Telephone band)</td><td rowspan="3">G.107 (Telephone band)
G.107 Appendix II
(Wideband)</td></tr>
<tr><td rowspan="2">Two way
(Conversational quality)</td><td colspan="2">FFS (Telephone-band)</td></tr>
<tr><td></td><td>P.561, P.562
(Telephone band)</td></tr>
<tr><td>Audio</td><td>One way
(Listening quality)</td><td>BS.1387-1</td><td></td><td></td></tr>
<tr><td>Video</td><td>One way
(Viewing quality)</td><td>J.144 (Cable TV)
FFS (Multimedia)</td><td>FFS (Cable TV)
FFS (Multimedia)</td><td></td></tr>
<tr><td rowspan="2">Speech/Audio and Video</td><td>One way</td><td rowspan="2" colspan="2">J.148 (Multimedia)</td><td>FFS (Video streaming)</td></tr>
<tr><td>Two way</td><td>G.1070
(Videophone)</td></tr>
<tr><td>Data</td><td>One way</td><td>FFS (Web browsing)</td><td></td><td>G.1030 Annex A
(Web browsing)</td></tr>
<tr><td colspan="5">NOTE – FFS stands for "for further study."</td></tr>
</table>

7 Framework

The framework of the opinion model treated in this Recommendation is illustrated in Figure 1. Its input parameters are video and speech quality parameters that are considered important in QoE/QoS planning. The model consists of three functions: video quality estimation, speech quality estimation, and multimedia quality integration functions. The degradation caused by pure delay is considered only in the multimedia quality integration function.

This Recommendation provides basic formulas for the above functions. The outputs from the model are multimedia quality (MM_q), video quality influenced by speech quality ($V_q(S_q)$), and speech quality influenced by video quality ($S_q(V_q)$).

The model assumes some specific evaluation conditions for terminals, environments, and evaluation contexts, and quality estimation under other evaluation conditions is currently under study. In that sense, these are the limitations of the current Recommendation.

It should be noted that the effects of a codec on subjective quality are dependent on its implementation. In particular, the quality of a video codec cannot be estimated based simply on the information about the coding technology (e.g., MPEG-4) used in the system under test, although specifying the implementation of a speech codec based on the standard (e.g., [ITU-T G.711]) employed in the system under test is relatively easy. For example, there are a number of different implementations for MPEG-4 codecs due to variations in coding-parameter settings and decoder characteristics. Therefore, the coefficients of video and speech quality estimation functions in this

Recommendation were determined by referring to tables prepared in advance for each video and speech codec. A coefficient database for video is illustrated in Figure 2. Such tables can be constructed by using the methodology provided in Annex A for video and by using [ITU-T P.833] or [ITU-T P.834] for speech.

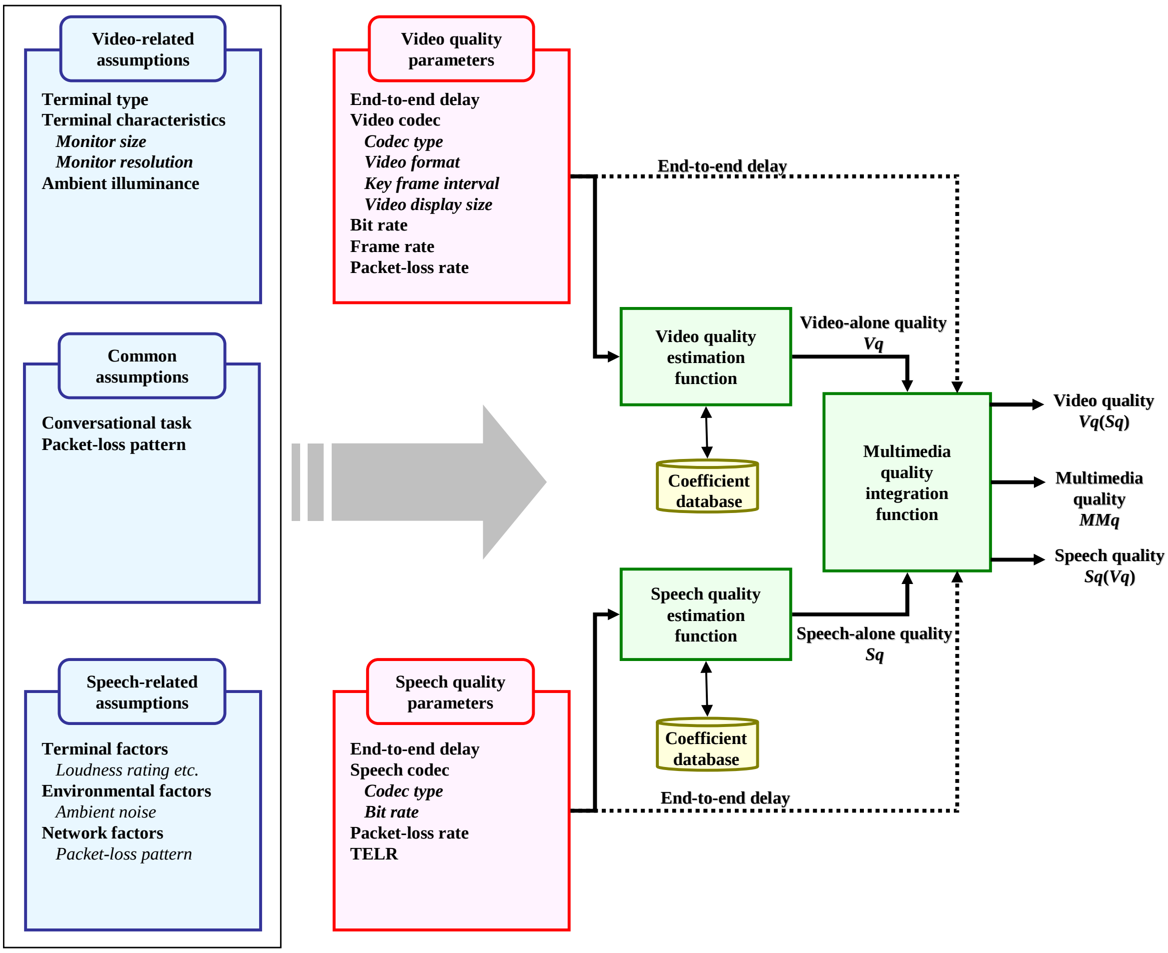

Figure 1 – Framework of multimedia communication quality assessment model

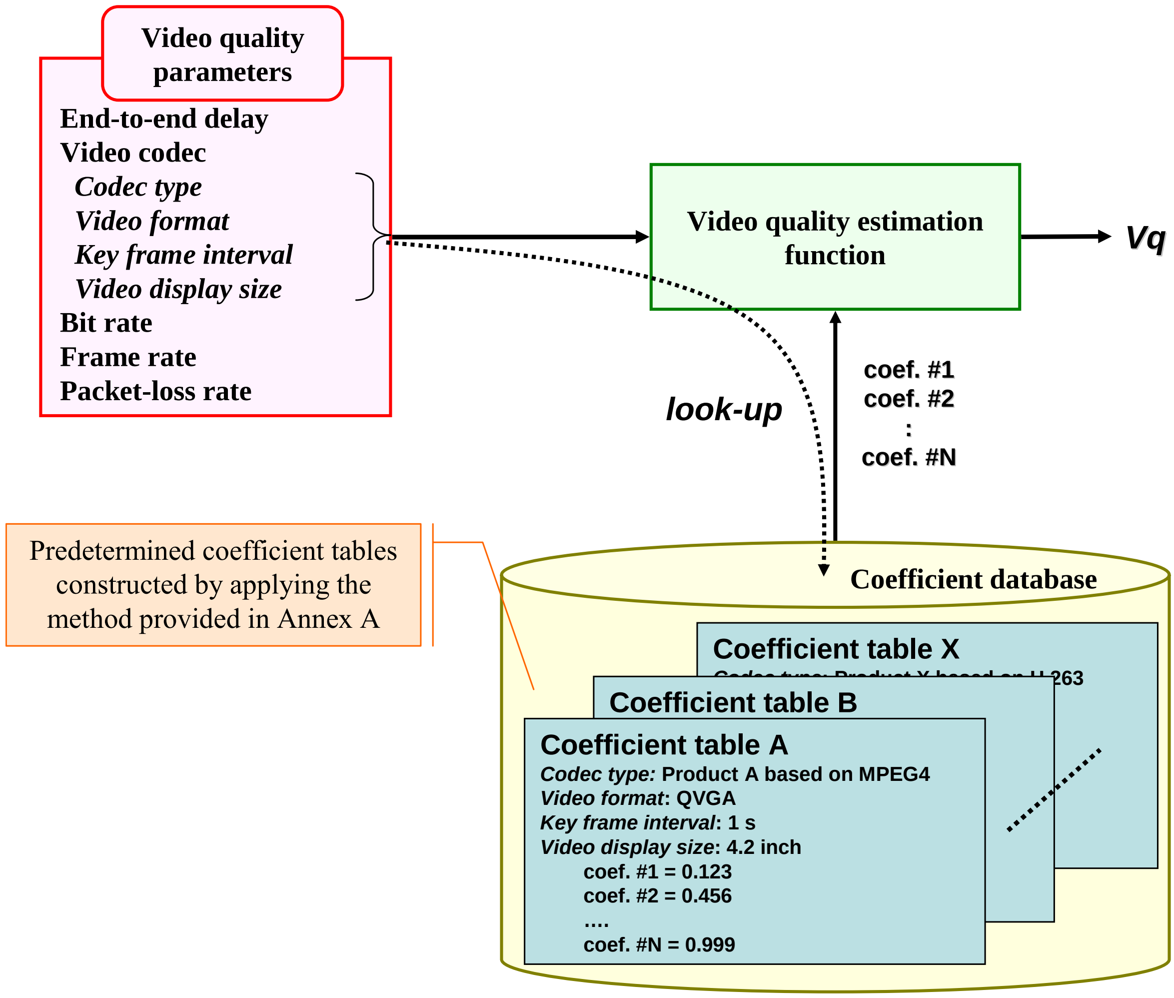

Figure 2 – Determination of coefficients that depend on codec implementation

8 Model assumptions

This clause describes conditions that the model assumes for terminals, environments, and evaluation contexts.

8.1 Speech-related assumptions

8.1.1 Terminal factors

The model assumes the use of a handset as the interface for the voice path. We recognize that in many cases the interface will be a hands-free or headset unit; further work is needed to include these in the G.107 model before they can be included as options for the multimedia model. This means that one needs to be very careful when he/she evaluates a mobile terminal, which usually has a hands-free function.

If a noise canceller and/or an automatic gain controller is applied, it is assumed that the device works without causing any additional degradation of speech signals.

8.1.2 Environmental factors

The assumed ambient noise is Hoth noise at 35 dB(A). Although other ambient noise conditions can be assumed to exist, especially in mobile applications, dealing with such conditions is for further study.

8.2 Video-related assumptions

8.2.1 Terminal factors

The model estimates video quality when that is evaluated by using a monitor whose specifications are listed in Table 3. Although this Recommendation treats video-telephony services using dedicated videophone terminals, desktop PCs, laptop PCs, PDAs, and mobile phones, the effects of terminal characteristics with specifications other than those shown in Table 3 are still under study.

NOTE – Specifications of monitors used in most PDAs and mobile phones are lower than those in Table 3. Thus, the model's predictions are often more sensitive than the users' opinions. Table 3 suggests nominal values, which are assumed by the model in this Recommendation, for use in QoE/QoS planning.

Table 3 – Assumptions about monitor characteristics

Monitor specifications	Value
Diagonal length[a)]	2-10 inches
Dot pitch	<0.30
Colour temperature	6500 K
Bit depth	8 bits/colour
Refresh rate	≥60 Hz
Brightness	100-300 cd/m^2
a) "Diagonal length" refers to the image size on the monitor screen.	

8.2.2 Environmental factors

The nominal ambient illumination is taken to be 500 lux. The model assumes video content to be a so-called "bust shot" with grey background, but that does not consider moving backgrounds in mobile applications, nor camera shaking due to hand movement.

8.3 Task-related assumptions

Conversational tasks in subjective quality evaluation affect the resultant conversational quality. In particular, the effects of delay have great impact on the interactivity in a conversation. Therefore, it is desirable that the model takes into account the kind of task assumed in the service. However, the only task currently considered in the model is "free conversation". This means that the video image of a conversation partner is displayed on the screen. Modelling of conversational tasks other than free conversation is for further study.

9 Model inputs

This clause describes the input parameters used in the model.

9.1 Speech quality parameters

The parameters described in this clause are similar to those in [ITU-T G.107]. Speech quality parameters not listed in this clause are assumed to take their default values as defined in [ITU-T G.107].

9.1.1 Speech delay (T_S [ms])

This refers to end-to-end, one-way delay in speech. Considering the delay in terminals, such as processing delay and jitter-buffer delay is extremely important. An input value of T_S must be less than 1000 ms.

9.1.2 Speech coding distortion (Ie_S)

Distortion due to speech coding needs to be quantified as Ie_S. The Ie_S values for ITU-T standard codecs are provided in Appendix I of [ITU-T G.113]. The Ie_S values for other codecs should be derived by applying the methods defined in [ITU-T P.833] or [ITU-T P.834].

9.1.3 Speech packet-loss robustness (Bpl_S)

Packet-loss robustness of a speech codec should be quantified as Bpl_S. The Bpl_S values for ITU-T standard codecs are provided in Appendix I of [ITU-T G.113].

9.1.4 Speech packet-loss rate (Ppl_S [%])

This refers to the end-to-end IP packet-loss rate in speech. Considering the packet loss in a terminal-jitter buffer and the packet loss in networks is extremely important. The value should be less than 20 [%].

9.1.5 Talker echo loudness rating (*TELR*)

This is the sum of SLR, RLR, and ERL in the talker-echo path. SLR and RLR are defined in [ITU-T P.79], and ERL is defined in [ITU-T G.122].

9.2 Video quality parameters

9.2.1 Video delay (T_V [ms])

This refers to end-to-end one-way delay in video. Considering the delay in terminals, such as processing delay and jitter-buffer delay is extremely important. An input value of T_V must be less than 1000 ms.

9.2.2 Video codec specifications

The model's coefficients (see Figure 2) for coding and packet-loss distortion are determined by looking up the coefficient database that is provided in Appendix I. For conditions other than those in Appendix I, one needs to derive coefficients by applying the method described in Annex A.

9.2.2.1 Codec type and implementation

This information is used to identify the specific implementation of a video codec under evaluation so that the model utilizes the coefficients appropriate to that implementation. The method to derive coefficients is provided in Annex A.

9.2.2.2 Spatial resolution

This parameter refers not to the actual/effective spatial resolution reflecting the performance of a camera and/or a display but to the theoretical spatial resolution employed in a codec. It is better to measure the effective spatial resolution, if possible, and reflect it in the quality estimation. [ITU-T P.800] provides a methodology for measuring the effective spatial resolution. However, how to reflect such results in the quality estimation model is still under study.

The model handles video whose size is between QQVGA and VGA.

9.2.2.3 Key-frame interval

This is the time interval in which video is coded solely from intra-frame information. This affects the effectiveness of video coding (i.e., quality vs video bit rate) and the robustness against packet-loss degradation.

9.2.3 Video packet-loss rate (Ppl_V [%])

This refers to end-to-end IP packet-loss rate in video. Considering the packet loss in a terminal-jitter buffer and the packet loss in networks is extremely important. The value should be less than 10 [%].

9.2.4 Video frame rate (Fr_V [fps])

This refers to the frame rate used in an encoder and does not reflect frame repetition used at a decoder, for example, in the case of packet loss. This Recommendation assumes that the range of the frame rate is from 1 to 30 fps.

9.2.5 Video bit rate (Br_V [kbit/s])

This refers to the video bit rate at an encoder.

10 Model outputs

The model outputs are the multimedia quality (MM_q), speech quality accounting for the influence of video quality ($S_q(V_q)$), and video quality accounting for the influence of speech quality ($V_q(S_q)$).

NOTE – The determination of $V_q(S_q)$ and $S_q(V_q)$ is for further study.

11 Model description

11.1 Speech quality estimation function

First, the speech quality parameters defined in clause 9.1 are mapped to a quality index Q as follows:

$$Q = 93.193 - Idte - Ie\text{-}eff \tag{11-1}$$

NOTE 1 – The quality index Q is equivalent to the transmission rating factor R defined in [ITU-T G.107], but the definition in this Recommendation is simplified due to the smaller number of input parameters.

NOTE 2 – Quality evaluation characteristics in multimedia applications might be different from those expected in telephony applications. Therefore, the model described in this clause is a provisional method. The applicability of [ITU-T G.107] in such applications is still under study.

NOTE 3 – The delay quality is considered separately in the multimedia quality integration function (see Figure 1), so equation 11-1 excludes *Idd*, which represents the degradation caused by pure delay in [ITU-T G.107].

Idte represents the degradation caused by talker echo and is defined as:

$$Idte = \left[\frac{94.769 - Re}{2} + \sqrt{\frac{(94.769 - Re)^2}{4} + 100} - 1 \right] \left(1 - \mathrm{e}^{-T_S} \right) \tag{11-2}$$

where:

$$\mathrm{Re} = 80 + 2.5(TERV - 14) \tag{11-3}$$

and:

$$TERV = TELR - 40 \log \frac{1 + \frac{T_S}{10}}{1 + \frac{T_S}{150}} + 6\mathrm{e}^{-0.3 T_S^{\,2}} \tag{11-4}$$

Ie-eff represents the degradation caused by speech coding, and packet loss and is defined as:

$$Ie\text{-}eff = Ie_S + (95 - Ie_S) \times \frac{Ppl_S}{Ppl_S + Bpl_S} \tag{11-5}$$

Speech quality S_q is defined as a function of the quality index Q.

For $Q < 0$: $S_q = 1$

For $0 < Q < 100$: $S_q = 1 + 0.035Q + Q(Q-60)(100-Q)7\times10^{-6}$ (11-6)

For $Q > 100$: $S_q = 4.5$

11.2 Video quality estimation function

11.2.1 Calculation of video quality, V_q

Video quality V_q is calculated using the video quality parameters defined in clause 9.2. V_q is expressed as:

$$V_q = 1 + I_{coding}\exp\left(-\frac{Ppl_V}{D_{PplV}}\right) \quad (11\text{-}7)$$

where I_{coding} represents the basic video quality affected by the coding distortion under a combination of video bit rate (Br_V [kbit/s]) and video frame rate (Fr_V [fps]), and the packet loss robustness factor D_{PplV} expresses the degree of video quality robustness due to packet loss where Ppl_V [%] represents the packet-loss rate.

11.2.2 Basic video quality affected by coding distortion, I_{coding}

The basic video quality affected by coding distortion I_{coding} is expressed as:

$$I_{coding} = I_{Ofr}\exp\left\{-\frac{\left(\ln(Fr_V)-\ln(O_{fr})\right)^2}{2D_{FrV}^{\ 2}}\right\} \quad (11\text{-}8)$$

The O_{fr} is an optimal frame rate that maximizes the video quality at each video bit rate (Br_V) and is expressed as:

$$O_{fr} = v_1 + v_2\ Br_V,\quad 1 \le O_{fr} \le 30,\ v_1 \text{ and } v_2\text{: const.} \quad (11\text{-}9)$$

where $Fr_V = O_{fr}$, $I_{coding} = I_{Ofr}$, I_{Ofr} represents the maximum video quality at each video bit rate (Br_V) and is expressed as:

$$I_{Ofr} = v_3 - \frac{v_3}{1+\left(\frac{Br_V}{v_4}\right)^{v_5}},\quad 0 \le I_{Ofr} \le 4,\ v_3,\ v_4, \text{ and } v_5\text{: const.} \quad (11\text{-}10)$$

D_{FrV} represents the degree of video quality robustness due to frame rate (Fr_V) and is expressed as:

$$D_{FrV} = v_6 + v_7\ Br_V,\quad 0 < D_{FrV},\ v_6 \text{ and } v_7\text{: const.} \quad (11\text{-}11)$$

Coefficients v_1, v_2, ..., and v_7 are dependent on codec type, video format, key frame interval, and video display size (see Appendix I).

11.2.3 Packet loss robustness factor, D_{PplV}

The packet loss robustness factor D_{PplV} represents the degree of video quality robustness against packet loss and is expressed as:

$$D_{PplV} = v_{10} + v_{11}\exp\left(-\frac{Fr_V}{v_8}\right) + v_{12}\exp\left(-\frac{Br_V}{v_9}\right),\quad 0 < D_{PplV} \quad (11\text{-}12)$$

where Ppl_V represents the packet-loss rate.

Coefficients v_8, v_9, ..., and v_{12} are dependent on codec type, video format, key frame interval, and video display size (see Appendix I).

11.3 Multimedia quality integration function

11.3.1 Calculation of the multimedia quality, MM_q

The multimedia quality MM_q is calculated using speech quality S_q, video quality V_q, speech delay T_S, and video delay T_V. MM_q is expressed as:

$$MM_q = m_1\, MM_{SV} + m_2\, MM_T + m_3\, MM_{SV}\, MM_T + m_4 \tag{11-13}$$

where MM_{SV} represents audiovisual quality, MM_T represents the audiovisual delay impairment factor, and coefficients m_1, m_2, ..., and m_4 are dependent on video display size and conversational task. MM_q is bounded between 1 and 5.

11.3.2 Audiovisual quality, MM_{SV}

The audiovisual quality MM_{SV} is expressed as:

$$MM_{SV} = m_5\, S_q + m_6\, V_q + m_7\, S_q\, V_q + m_8 \tag{11-14}$$

Coefficients m_5, m_6, ..., and m_8 are dependent on video display size and conversational task. MM_{SV} is bounded between 1 and 5.

11.3.3 Audiovisual delay impairment factor, MM_T

The MM_T represents the degree of the audiovisual quality degradation due to audiovisual delay and synchronization and is expressed as:

$$MM_T = \max\{AD + MS, 1\} \tag{11-15}$$

$$AD = m_9\,(T_S + T_V) + m_{10} \tag{11-16}$$

$$MS = \min\{m_{11}\,(T_S - T_V) + m_{12}, 0\} \quad if\ T_S \geq T_V \tag{11-17}$$

and:

$$MS = \min\{m_{13}\,(T_V - T_S) + m_{14}, 0\} \quad if\ T_S < T_V \tag{11-18}$$

where AD represents the absolute audiovisual delay and MS represents the audiovisual media synchronization.

Coefficients m_9, m_{10}, ..., and m_{14} are dependent on video display size and conversational task.

NOTE 1 – Provisional values of m_i were developed for two different video displays, viz. a 4.2-video display size with a QVGA video format and a 2.1-inch video display with a QQVGA video format. The conversational task was free conversation. These values are provided in Appendix II.

NOTE 2 – Currently, the derivation of $V_q(S_q)$ and $S_q(V_q)$, which are video quality affected by speech quality and vice versa, is under study.

12 Accuracy of model

The Pearson product-moment correlation between the empirical subjective quality and the quality estimate generated by the model was used for evaluate the accuracies of the speech quality estimation function, the video quality estimation function, and the multimedia quality integration function. The correlation r should be calculated over all the test data sets as follows:

$$r = \frac{\sum_{v=1}^{V} \left(y_v - \bar{y}\right)\left(x_v - \bar{x}\right)}{\sqrt{\sum_{v=1}^{V} \left(y_v - \bar{y}\right)^2 \sum_{v=1}^{V} \left(x_v - \bar{x}\right)^2}}$$

where V is the number of test data sets, and the mean values of the data sets are calculated as follows:

$$\bar{x} = \frac{1}{V} \sum_{v=1}^{V} x_v$$

and:

$$\bar{y} = \frac{1}{V} \sum_{v=1}^{V} y_v$$

where:

x_v represents the estimated quality of test data; and

y_v represents the subjective quality of test data.

The accuracy of the speech quality estimation function, which is presented in [ITU-T G.107], is described in [b-Möller].

The accuracy of the video quality estimation function was verified by the following manner:

1) The validity of the forms (equations 11-7 to 11-12) was verified by using the subjective quality database employing H.264 and MPEG-4 codecs (DB#1 to DB#4 in Table 4). Here, the coefficients v_1, v_2, ..., and v_{12} were optimized for each database. The cross-correlation is about 0.975 on average [b-Yamagishi 1], [b-Yamagishi 2].

2) The validity of the optimized coefficients v_1, v_2, ..., and v_{12} was verified by applying them to unknown data. This was done by optimizing the coefficients by DB#5 and DB#7, and applying them to DB#6 and DB#8 (see Table 5). The cross-correlation is about 0.955 on average [b-ITU-T COM12-C34].

Table 4 – Verification of form of video quality estimation function for various terminals

	r	**CT**	**VF**	**KFI [s]**	**VDS [inches]**
DB#1	0.967 [b-Yamagishi 1]	H.264	QVGA	1	4.2
DB#2	0.987 [b-Yamagishi 1], [b-Yamagishi 2]	MPEG-4	QVGA	1	4.2
DB#3	0.973 [b-Yamagishi 2]	MPEG-4	QVGA	1	8.5
DB#4	0.972 [b-Yamagishi 2]	MPEG-4	VGA	1	8.5

Table 5 – Accuracy of video quality estimation function

	r	**CT**	**VF**	**KFI [s]**	**VDS [inches]**
DB#5	0.951 [b-ITU-T COM12-C34]	MPEG-4	QVGA	1	4.2
DB#6	0.961 [b-ITU-T COM12-C34]	MPEG-4	QVGA	1	4.2
DB#7	0.958 [b-ITU-T COM12-C34]	MPEG-4	QQVGA	1	2.1
DB#8	0.949 [b-ITU-T COM12-C34]	MPEG-4	QQVGA	1	2.1

The accuracy of the multimedia quality integration function was verified as follows.

The speech and video quality (S_q and V_q) were estimated by the speech and video quality estimation functions described in clauses 11.1 and 11.2, respectively, in this Recommendation. Next, we fed these values and speech and video delay values (T_S and T_V) into the multimedia quality integration function given in clause 11.3. By comparing the estimated multimedia quality with multimedia quality obtained by a conversational subjective test, we evaluated the validity of the multimedia quality integration function.

We used a G.711 codec without a packet-loss concealment algorithm as a speech codec. The speech packet-loss rate varied from 0 to 10%. No echo was introduced in the evaluation system. The video codec was MPEG-4, and the video bit rate was 2 Mbit/s and 1 Mbit/s for QVGA and QQVGA, respectively. The video packet-loss rate varied from 0 to 5%. The video frame rate was between 2 to 30 fps. One-way delay varied from 167 to 1000 ms, and was controlled for speech and video separately to evaluate the effects of lip synchronization. The number of experimental conditions was 88.

The subjective testing method was 5-grade conversational ACR defined in [ITU-T P.910]. There were 32 subjects. The viewing distances were 50 and 80 cm for QQVGA and QVGA, respectively. The number of judgments by each subject was 140. All the conditions (88) were assessed by each subject. Some were assessed twice or more.

The estimation accuracy of the G.1070 model, including speech and video quality estimation functions, is demonstrated in Table 6.

NOTE – These databases were used for optimizing the coefficients of the multimedia quality integration function under each VDS condition. The coefficients are provided in Appendix II.

Table 6 – Accuracy of multimedia communication quality assessment model

	r	**Speech codec**	**Video codec**	**VF**	**KFI [s]**	**VDS [inches]**
DB#9	0.83	G.711	MPEG-4 at 2 Mbit/s	QVGA	1	4.2
DB#10	0.91	G.711	MPEG-4 at 1 Mbit/s	QQVGA	1	2.1

Annex A

Methodology for deriving coefficients in video quality estimation function with respect to coding and packet-loss degradations

(This annex forms an integral part of this Recommendation)

A.1 Methodology for deriving coefficients v_1, v_2, ..., and v_7

Using the subjective video quality MOS, which is called V_{qs} hereafter, for various video bit rate (Br_V) and video frame-rate (Fr_V) conditions, coefficients v_1, v_2, ..., and v_7 are calculated in the following four steps.

Step A.1.1: Calculation of values I_{Ofr}, O_{fr}, and D_{fr}

1) If we employ M different frame rates for each video bit-rate condition b_n, we obtain Table A.1.

Table A.1 – Relationships among Br_V, Fr_V, and V_q

Br_V	Fr_V	V_q
b_n	f_1	$V_{qs}(b_n,f_1)$
b_n	f_2	$V_{qs}(b_n,f_2)$
...	...	...
b_n	f_m	$V_{qs}(b_n,f_m)$
...	...	...
b_n	f_M	$V_{qs}(b_n,f_M)$

NOTE 1 – M represents the number of frame rate conditions. $f_1 > f_2 > ..... > f_M$.

NOTE 2 – $V_{qs}(b_n, f_m)$ represents the MOS under the condition with a video bit rate of b_n and a frame rate of f_m.

2) By applying the data set in Table A.1 to equation 11-7, Ofr, I_{Ofr}, and D_{fr} are approximated for each video bit rate b_n based on the least square approximation (LSA). As a result, we can obtain Table A.2.

Table A.2 – Relationship between Br_V, I_{Ofr}, O_{fr}, and D_{fr}

Br_V	O_{fr}	I_{Ofr}	D_{fr}
b_1	O_1	I_1	D_1
b_2	O_2	I_2	D_2
...	...	...	...
b_n	O_n	I_n	D_n
...	...	...	...
b_N	O_N	I_N	D_N

NOTE 3 – N represents the number of video bit-rate conditions. $b_1 > b_2 > ..... > b_N$.

Step A.1.2: Calculation of coefficients v_1 and v_2

By applying b_n and O_n for n = 1, 2, ..., N in Table A.2 to equation 11-9, coefficients v_1 and v_2 are approximated based on the LSA.

Step A.1.3: Calculation of coefficients v_3, v_4, and v_5

By applying b_n and I_n for n = 1, 2, ..., N in Table A.2 to equation 11-10, coefficients v_3, v_4, and v_5 are approximated based on the LSA.

Step A.1.4: Calculation of coefficients v_6 and v_7

By applying b_n and D_n for n = 1, 2, ..., N to equation 11-11, coefficients v_6 and v_7 are approximated based on the LSA.

A.2 Methodology for deriving coefficients v_8, v_9, …, and v_{12}

Using the subjective video quality (V_{qs}) related to video bit rate (Br_V), video frame rate (Fr_V), and video packet-loss rate (Ppl_V), coefficients v_8, v_9, …, and v_{12} are calculated in the following four steps.

NOTE 1 – The subjective quality characteristics of packet-loss degradation often depend on the duration of video sequences used in a subjective test, so one should use video sequences that have a reasonable length (e.g., 1 min).

Step A.2.1: Calculation of values D_{PplV}

By applying I_{coding}, which is calculated by using the coefficients derived in clause A.1, and subjective video quality (V_{qs}) to equation 11-7, the packet loss robustness factor D_{PplV} is approximated based on the LSA for each combination of Br_V and Fr_V, as shown in Table A.3.

Table A.3 – Relationships among video bit rate, video frame rate, and D_{Pplv}

D_{PplV}		Fr_V					
		f_1	f_2		f_m		f_M
Br_V	b_1	D_{b1f1}	D_{b1f2}	...	...	...	D_{b1fM}
	b_2	D_{b2f1}	D_{b2f2}	...	...	...	D_{b2fM}
	...	...	...	...	...	...	...
	b_n	D_{bnf1}	D_{bnf2}	...	D_{bnfm}	...	D_{bnfM}
	...	...	...	...	...	...	...
	b_N	D_{bNf1}	D_{bNf2}	...	...	...	D_{bNfM}

NOTE 1 – N represents the number of video bit-rate conditions.

NOTE 2 – M represents the number of video frame-rate conditions.

NOTE 3 – D_{bnfm} indicates a temporary value of the packet loss robustness factor D_{PplV} for a video bit rate of b_n and a frame rate of f_m.

Step A.2.2: Calculation of coefficient v_8

By applying f_m and $D_{PplV} = D_{b1fm}$ for m = 1, 2, ..., M to equation A-1, coefficients a, b, and v_8 are approximated based on the LSA.

$$D_{PplV} = a + b \exp\left(-\frac{Fr_V}{v_8}\right) \quad \text{(A-1)}$$

Step A.2.3: Calculation of coefficient v_9

By applying b_n and $D_{PplV} = D_{bnf1}$, for n = 1, 2, ..., N to equation A-2, coefficients c, d, and v_9 are approximated based on the LSA.

$$D_{PplV} = c + d \exp\left(-\frac{Br_V}{v_9} \right) \tag{A-2}$$

NOTE 2 – Coefficients a, b, c, and d are temporary and never used in the following calculation.

Step A.2.4: Calculation of coefficients v_{10}, v_{11}, and v_{12}

By applying v_8, v_9, $D_{PplV} = D_{bnfm}$, $Br_V = b_n$, and $Fr_V = f_m$ for n = 1, 2, ..., N and m = 1, 2, ..., M to equation 11-12, coefficients v_{10}, v_{11}, and v_{12} are approximated based on LSA.

Appendix I

Coefficients in video quality estimation function with respect to coding and packet-loss degradations

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

This appendix provides the provisional coefficient tables to be used for the video quality estimation. Table I.1 summarizes the conditions under which each coefficient table was constructed.

NOTE 1 – The provisional coefficient tables given in this appendix cannot be applied to arbitrary MPEG-4 codecs. That is dependent on the implementation and setting of codec, as noted in clause 7. Therefore, if one needs coefficient values for a codec which is not included in this table, the procedure described in Annex A should be used to create appropriated tables.

Table I.1 – Conditions for deriving coefficient tables

Factors	# 1	# 2
Codec type	MPEG-4	MPEG-4
Video format	QVGA	QQVGA
Key frame interval (s)	1	1
Video display size (inch)	4.2	2.1

The resultant provisional coefficient values are provided in Table I.2.

NOTE 2 – These provisional coefficient values were determined based on subjective tests with video sequences 10 s. Therefore, the quality estimation based on these coefficients may result in optimistic evaluation in comparison with that of the video quality of longer video sequences in evaluating the effects of packet loss.

Table I.2 – Provisional coefficient table for video quality estimation function

Coefficients	# 1	# 2
v_1	1.431	7.160
v_2	2.228×10^{-2}	2.215×10^{-2}
v_3	3.759	3.461
v_4	184.1	111.9
v_5	1.161	2.091
v_6	1.446	1.382
v_7	3.881×10^{-4}	5.881×10^{-4}
v_8	2.116	0.8401
v_9	467.4	113.9
v_{10}	2.736	6.047
v_{11}	15.28	46.87
v_{12}	4.170	10.87

Appendix II

Coefficients in multimedia quality integration function

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

This appendix provides the provisional coefficient values for the multimedia quality integration function. As stated in clause 11.3, the coefficients are dependent on the video display size and conversational task. The coefficient tables in this appendix assume two different video display sizes, which are 4.2, and 2.1 [inch]. They were derived by using "free conversation" as a conversational task.

Table II.1 – Provisional coefficients of multimedia quality integration function

Coefficients	4.2 inch	2.1 inch
m_1	-4.457×10^{-1}	-6.966×10^{-1}
m_2	-6.638×10^{-1}	-8.127×10^{-1}
m_3	4.042×10^{-1}	4.562×10^{-1}
m_4	2.321	3.003
m_5	-3.255×10^{-1}	-1.638×10^{-1}
m_6	3.309×10^{-1}	3.626×10^{-1}
m_7	1.494×10^{-1}	1.291×10^{-1}
m_8	5.457×10^{-1}	5.456×10^{-1}
m_9	-3.235×10^{-4}	-1.251×10^{-4}
m_{10}	3.915	3.763
m_{11}	-1.377×10^{-3}	-1.065×10^{-3}
m_{12}	0.000	1.465×10^{-2}
m_{13}	-1.095×10^{-3}	-1.002×10^{-3}
m_{14}	0.000	0.000

Bibliography

[b-Bijl] BIJL (P.), VALETON (J.M.): Guidelines for accurate TOD measurement, *Proc. SPIE. 3701*, pp.14-25, 1999.

[b-Möller] MÖLLER (S.): Assessment and Prediction of Speech Quality in Telecommunications, *Kluwer Academic Publishers*. Boston, MA, 2000.

[b-Yamagishi 1] YAMAGISHI (K.), HAYASHI (T.): Verification of Video Quality Opinion Model for Videophone Services, *2nd ISCA Tutorial & Research Workshop on Perceptual Quality of Systems*, pp.143-148, Sept. 2006.

[b-Yamagishi 2] YAMAGISHI (K.), HAYASHI (T.): Opinion Model for Estimating Video Quality of Videophone Services, IEEE GLOBECOM 2006, QRP08-1, Nov. 2006.

[b-ITU-T COM12-C34] ITU-T Contribution COM 12-C 34: *Verification Of Video Quality Estimation Model For Videophone Services (NTT)*, Jan. 2007.

SERIES OF ITU-T RECOMMENDATIONS

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

Printed in Switzerland
Geneva, 2007