International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

G.1091

(10/2014)

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

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

Quality of Experience requirements for telepresence services

Recommendation ITU-T G.1091

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
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.1091

Quality of Experience requirements for telepresence services

Summary

Telepresence is an interactive audio-visual communications experience between remote locations, where users enjoy a strong sense of realism and presence between participants by optimizing a variety of attributes such as audio and video quality, eye contact, gaze awareness, body language, spatial audio, coordinated environments and natural image size. In this regard, delivering a high-level of quality of experience (QoE) is an important factor. Recommendation ITU-T G.1091 provides guidelines for achieving high QoE in telepresence services. First, a high-level description of a telepresence system is described, with an overview of important use cases. Then, specific audio and visual parameters, and other characteristics that are important from a QoE perspective are defined. Where possible, specific values for the parameters necessary to achieve high QoE are recommended, along with references to associated test methodologies.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T G.1091	2014-10-29	12	11.1002/1000/12310

Keywords

QoE, telepresence.

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

FOREWORD

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

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

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

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

NOTE

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

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

INTELLECTUAL PROPERTY RIGHTS

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

As of the date of approval of this Recommendation, ITU had not received notice of intellectual property, protected by patents, 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/.

Table of Contents

		Page
1	Scope	1
2	References	1
3	Definitions	1
	3.1 Terms defined elsewhere	1
	3.2 Terms defined in this Recommendation	2
4	Abbreviations and acronyms	2
5	Conventions	2
	5.1 Requirement terminology	2
	5.2 Icons used	2
6	High-level QoE considerations for telepresence services	3
	6.1 Telepresence system description	4
	6.2 Use cases	4
7	Audio characteristics	5
	7.1 Send and receive loudness ratings	5
	7.2 Send and receive side frequency responses	5
	7.3 Send and receive side distortion	6
	7.4 Send and receive side noise	6
	7.5 Directionality of audio (spatial characteristics)	6
	7.6 Acoustic echo/double talk	6
	7.7 Intelligibility of audio	6
8	Video characteristics	6
	8.1 Actual size	6
	8.2 Resolution (definition)	6
	8.3 Image quality	6
	8.4 Eye gaze	7
	8.5 Spatial considerations for video	7
9	Display characteristics	7
	9.1 Screen size	7
	9.2 Aspect ratios	7
	9.3 Continuity across borders	7
	9.4 Colour space (4:2:2, 4:2:0, 4:4:4)	7
	9.5 Colour fidelity	7
	9.6 Mirrored image	7
10	Audiovisual characteristics	7
	10.1 Absolute delay of audio and video streams	8
	10.2 Lip sync	8
11	Room characteristics	8

Page

11.1 Background noise 8
11.2 Reverberation 8
11.3 Lighting 8
11.4 Sound uniformity 8
11.5 Impact of seating arrangements 8
11.6 Temperature 8
11.7 Humidity 8
11.8 Audibility in local room 8
12 Other characteristics 9
12.1 Ease of use 9
Bibliography 10

Recommendation ITU-T G.1091

Quality of Experience requirements for telepresence services

1 Scope

This Recommendation provides guidelines for achieving high quality of experience (QoE) in telepresence services. First, a high-level description of a telepresence system is described, with an overview of important use cases. Then, specific audio and visual parameters, and other characteristics that are important from a QoE perspective are defined. Where possible, specific values for the parameters necessary to achieve high QoE are recommended, along with references to associated test methodologies. The list of parameters and requirements is not to be considered as an exhaustive list, and it is expected that this Recommendation will be updated as more experience is gained with telepresence systems.

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 F.734] Recommendation ITU-T F.734 (2014), *Definitions, requirements and use cases for telepresence systems*.

[ITU-T G.131] Recommendation ITU-T G.131 (2003), *Talker echo and its control*.

[ITU-T H.420] Recommendation ITU-T H.420 (2014), *Telepresence system architecture*.

[ITU-T P.300] Recommendation ITU-T P.300 (2001), *Transmission performance of group audio terminals (GATs).*

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

3.1.1 actual size [ITU-T F.734]: A rendered figure in a display is the same size as if the person is in the room.

3.1.2 gaze awareness [ITU-T F.734]: Awareness of gaze direction of persons by watching their eyes, head and body position. There is awareness of both direct gaze and averted gaze. Direct gaze is the perception of persons looking at each other directly and simultaneously. Averted gaze is the perception that other persons are looking at someone/something else in the environment.

3.1.3 gaze direction [ITU-T F.734]: The vector positioned along the visual axis, pointing from the fovea of the looker through the centre of the pupil to the gazed-at spot.

3.1.4 eye contact [ITU-T F.734]: Technically equivalent to a mutual direct-gaze between two persons communicating.

3.1.5 render [b-IETF RFC 7262]: The process of generating a representation from a media, such as displayed motion video or sound emitted from loudspeakers.

3.1.6 telepresence [ITU-T F.734]: An interactive audio-visual communications experience between remote locations, where the users enjoy a strong sense of realism and presence between participants by optimizing a variety of attributes such as audio and video quality, eye contact, gaze awareness, body language, spatial audio, coordinated environments and natural image size.

3.1.7 telepresence system [ITU-T H.420]: A set of functions, devices and network elements which are able to capture, deliver, manage and render multiple high quality interactive audio and video signals in a telepresence conference. An appropriate number of devices (e.g., cameras, screens, loudspeakers, microphones, codecs, multipoint control units, PCs) and environmental characteristics are used to establish telepresence.

3.2 Terms defined in this Recommendation

None.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

CRC Call and Resource Controller

GW Gateway

IP Internet Protocol

MCU Multipoint Control Unit

PC Personal Computer

PLMN Public Land Mobile Network

QoE Quality of Experience

5 Conventions

5.1 Requirement terminology

In this Recommendation the following conventions are used:

– "Shall" indicates a mandatory requirement.

– "Should" indicates a suggested but optional course of action.

– "May" indicates an optional course of action rather than a recommendation that something take place.

5.2 Icons used

Figure 1 shows the meaning of icons used in this Recommendation.

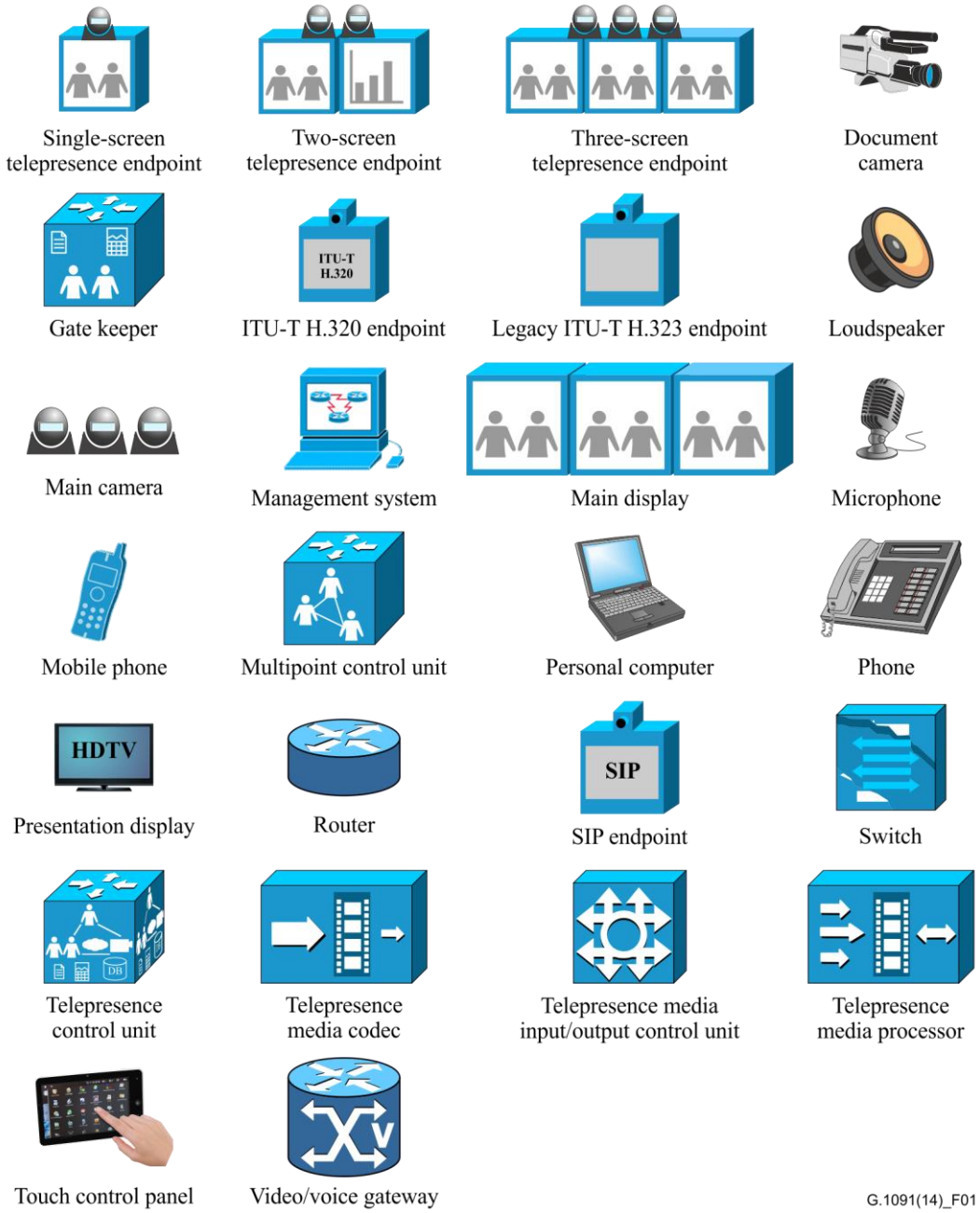

Figure 1 – Icons used in this Recommendation

6 High-level QoE considerations for telepresence services

This clause provides a high-level description of telepresence systems. Some scenarios and use cases are also provided to provide a context for assessing the important factors from the point of view of QoE.

6.1 Telepresence system description

The term "telepresence systems" is used to refer to conferencing systems that provide high definition and high quality audio/video, enabling a "being-there" experience. A telepresence system includes endpoints (both telepresence and legacy), multipoint control unit (MCU), call and resource controller (CRC), gateway (GW), management system, etc., as shown in Figure 2. [ITU-T H.420], "Telepresence System Architecture", provides a detailed description of a telepresence system.

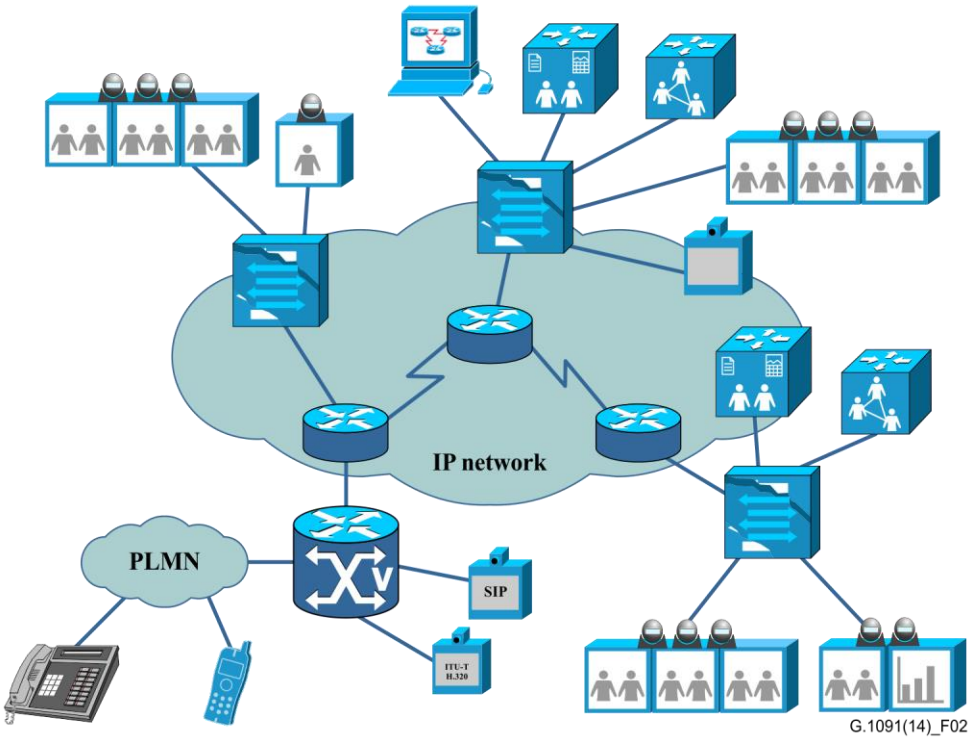

Figure 2 – High-level telepresence architecture network model

6.2 Use cases

Clause 6 of [ITU-T F.734], "Definitions, requirements, and use cases for telepresence systems", presents a set of use cases describing typical scenarios. The most common and basic use cases are introduced and illustrated in detail, which are described from the viewpoint of the users.

6.2.1 Use case 1: Symmetric point-to-point meeting

In this use case, each of the two sites has an identical number of screens, with cameras having fixed fields of view, and one camera for each screen. The number of screens or cameras is not relevant as long as both sites have the same number. The sound type is the same at each end. Figure 3 illustrates the use case.

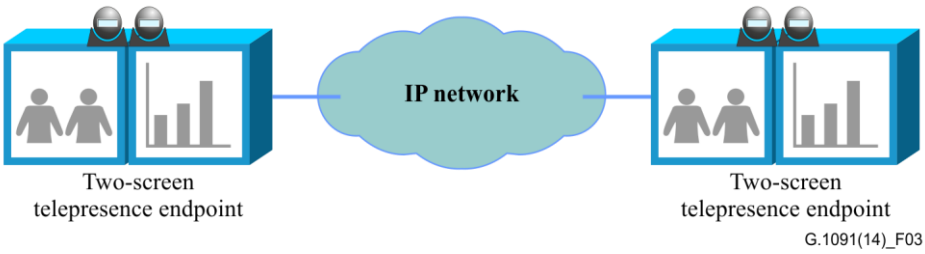

Figure 3 – Use case 1: Symmetric point-to-point meeting

For further information, see clause 3.1 of [b-IETF RFC 7205].

6.2.2 Use case 2: Multipoint meeting

In this use case of a multipoint telepresence conference, there are more than two sites participating. Figure 4 illustrates the use case. The number of screens or cameras can be different or the same. Each site can have different audio/video characteristics.

For further information, see clause 3.3 of [b-IETF RFC 7205].

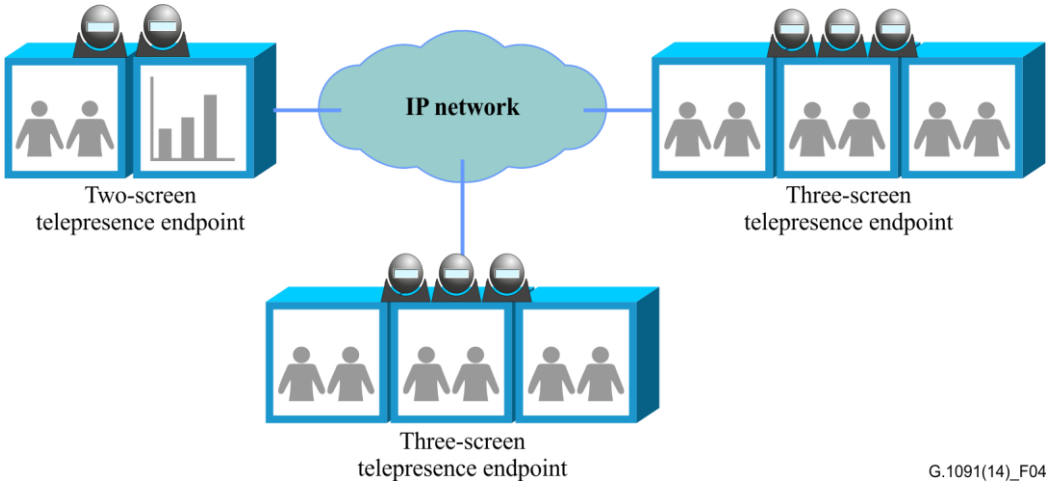

Figure 4 – Use case 2: Multipoint meeting

7 Audio characteristics

This clause describes the audio characteristics related to QoE of telepresence services in more detail.

7.1 Send and receive loudness ratings

Telepresence systems shall send audio to other telepresence systems with a SendLoudnessRating of (+ 13 – F_s) dB, measured in accordance with [ITU-T P.300].

In order to take into account the difference between the reference test positioning and the actual microphone-talker operating distance (d_s), for which the terminal is adjusted, the following correction factor F_s is defined:

$$F_s[\text{dB}] = 20\ \text{Log}\left(\frac{d_s}{0.5}\right) \qquad (d_s \text{ in metres})$$

Telepresence systems shall receive audio from other telepresence systems with a nominal Receive Loudness Rating of (+ 5 – F_r) dB, measured in accordance with [ITU-T P.300].

In order to take into account the difference between the reference test positioning and the actual loudspeaker-listener operating distance (d_r) for which the terminal is adjusted, the following correction factor F_r is defined:

$$F_r[\text{dB}] = 20\ \text{Log}\left(\frac{d_r}{0.5}\right) \qquad (d_r \text{ in metres})$$

7.2 Send and receive side frequency responses

Telepresence systems shall provide audio with send and receive side frequency responses from 100 Hz to 20 kHz.

7.3 Send and receive side distortion

Telepresence systems shall provide audio with send and receive side distortion less than 3%. Note, different specifications exist for send and receive.

For send side distortion the requirement is 2% (1% is a desired objective).

For received side distortion the requirement is 3%.

7.4 Send and receive side noise

Send side noise should be less than –63 dBm0p. Receive side noise should be less than 40 dBA (under maximum volume).

7.5 Directionality of audio (spatial characteristics)

In telepresence systems, all the seats shall sense directions (e.g., left, middle, right for three screen telepresence endpoint) for rendering.

The person who is talking shall be identified in correct spatial geometry.

7.6 Acoustic echo/double talk

Single talk: The weighted terminal coupling loss (TCLw) should be greater than or equal to 45 dB.

Double talk: Talker echo loss rating (TELR, as defined in [ITU-T G.131]) double talk should be greater than 27 dB.

7.7 Intelligibility of audio

Telepresence systems shall provide audio with intelligibility greater than 94%.

8 Video characteristics

This clause describes the video characteristics related to QoE of telepresence services in more detail.

8.1 Actual size

Telepresence systems shall provide an actual-sized video display as far as possible:

The difference in size between the original and displayed video should be limited within 10% of the original.

8.2 Resolution (definition)

Telepresence systems shall provide video with resolution higher than 720 P.

Telepresence systems shall provide a harmonic view among image displays of different resolutions.

(One example could be that multiple video captures of different resolutions are to be displayed in one room.)

8.3 Image quality

Telepresence systems shall provide images without noticeable geometric distortion.

The overall geometric distortion should be less than 10%.

Telepresence systems shall provide images without noticeable noise.

8.4 Eye gaze

The remote participants' video images should be rendered within a tolerable angle in both horizontal and vertical axes from the gaze direction between the two communicating participants. The tolerable angles should be 3-5 degrees horizontal and up to 12 degrees upward and 8 degrees downward.

8.5 Spatial considerations for video

Telepresence systems shall keep the spatial relationship of video as of physical relationship of participants.

Sync of videos in different screens should also be considered.

9 Display characteristics

This clause describes the display characteristics related to QoE of telepresence services in more detail. In contrast to the video characteristics, displays characteristics focus on physical arrangement for screens, consistency across multiple screens, etc.

9.1 Screen size

Ideally, all screens of a telepresence room should be of the same size. However, if different sizes are used in the same room, they should display images for different purposes.

9.2 Aspect ratios

Telepresence systems shall provide a harmonized view among screens of different aspect ratios.

The displayed image size should not differ by more than 5% from the original image size even when different aspect ratios are used.

9.3 Continuity across borders

Telepresence systems shall provide a continuous view across frame borders of adjacent screens.

This is related to clause 9.1 and also illustrated in Figure 10 in clause 7 of [ITU-T F.734].

The width of the discontinuous part should be less than 3 cm.

9.4 Colour space (4:2:2, 4:2:0, 4:4:4)

Telepresence systems shall provide a harmonized view among displays of different colour spaces.

9.5 Colour fidelity

Telepresence systems shall ensure unnoticeable change in displayed colour, especially for skin colour and white balancing.

9.6 Mirrored image

Telepresence systems shall not provide mirrored images of participants by default.

When the local images are displayed on the local screens, it is recommended to provide mirrored images for the participants.

10 Audiovisual characteristics

This clause describes the audiovisual characteristics related to QoE of telepresence services in more detail.

10.1 Absolute delay of audio and video streams

– For telepresence systems: the end-to-end video delay shall be less than or equal to 320 ms.

– End-to-end audio delay shall be less than or equal to 280 ms.

10.2 Lip sync

Audio streams should be less than 40 ms ahead of video streams, or less than 60 ms behind video streams.

11 Room characteristics

11.1 Background noise

Telepresence systems shall enjoy a comfortable environment with background noise less than 40 dBA.

11.2 Reverberation

Early reflecting sound should be between 1 kHz and 8 kHz and 15 ms after direct sound should attenuate 10 dB against direct sound.

11.3 Lighting

Telepresence systems shall provide a suitable lighting mechanism in the room, which includes but is not limited to, gamut, colour temperature, colour rendering index, type of illumination, etc.

Telepresence systems should provide a colour temperature of 6’500 K, a luminance of 250 Lux.

11.4 Sound uniformity

Telepresence systems shall provide a uniform sound effect (to within 3 dB) for users in the same room.

11.5 Impact of seating arrangements

Please note that not all people share the same QoE in the room.

The maximum QoE difference should be less than 20%.

11.6 Temperature

The room temperature should be set between 10~25 C, a suggested temperature is 18 C.

11.7 Humidity

The relative humidity should be between 40% and 60%.

11.8 Audibility in local room

Audibility for a single word should be greater than or equal to 90%.

12 Other characteristics

12.1 Ease of use

The user interface for frequently used functions shall be optimized. The number of pushes or touches should be no more than five.

A tutorial should be provided.

The set-up of a telepresence call shall not exceed seven seconds.

Bibliography

[b-IETF RFC 7205] IETF RFC 7205 (2014), *Use Cases for Telepresence Multistreams*.

[b-IETF RFC 7262] IETF RFC 7262 (2014), *Requirements for Telepresence Multistreams*.

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	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
Series Z	Languages and general software aspects for telecommunication systems

Printed in Switzerland
Geneva, 2015