International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

H.420

(10/2014)

SERIES H: AUDIOVISUAL AND MULTIMEDIA SYSTEMS

Infrastructure of audiovisual services – Telepresence

Telepresence system architecture

Recommendation ITU-T H.420

ITU-T H-SERIES RECOMMENDATIONS

AUDIOVISUAL AND MULTIMEDIA SYSTEMS

CHARACTERISTICS OF VISUAL TELEPHONE SYSTEMS	H.100–H.199
INFRASTRUCTURE OF AUDIOVISUAL SERVICES	
General	H.200–H.219
Transmission multiplexing and synchronization	H.220–H.229
Systems aspects	H.230–H.239
Communication procedures	H.240–H.259
Coding of moving video	H.260–H.279
Related systems aspects	H.280–H.299
Systems and terminal equipment for audiovisual services	H.300–H.349
Directory services architecture for audiovisual and multimedia services	H.350–H.359
Quality of service architecture for audiovisual and multimedia services	H.360–H.369
Telepresence	**H.420–H.429**
Supplementary services for multimedia	H.450–H.499
MOBILITY AND COLLABORATION PROCEDURES	
Overview of Mobility and Collaboration, definitions, protocols and procedures	H.500–H.509
Mobility for H-Series multimedia systems and services	H.510–H.519
Mobile multimedia collaboration applications and services	H.520–H.529
Security for mobile multimedia systems and services	H.530–H.539
Security for mobile multimedia collaboration applications and services	H.540–H.549
Mobility interworking procedures	H.550–H.559
Mobile multimedia collaboration inter-working procedures	H.560–H.569
BROADBAND, TRIPLE-PLAY AND ADVANCED MULTIMEDIA SERVICES	
Broadband multimedia services over VDSL	H.610–H.619
Advanced multimedia services and applications	H.620–H.629
Ubiquitous sensor network applications and Internet of Things	H.640–H.649
IPTV MULTIMEDIA SERVICES AND APPLICATIONS FOR IPTV	
General aspects	H.700–H.719
IPTV terminal devices	H.720–H.729
IPTV middleware	H.730–H.739
IPTV application event handling	H.740–H.749
IPTV metadata	H.750–H.759
IPTV multimedia application frameworks	H.760–H.769
IPTV service discovery up to consumption	H.770–H.779
Digital Signage	H.780–H.789
E-HEALTH MULTIMEDIA SERVICES AND APPLICATIONS	
Interoperability compliance testing of personal health systems (HRN, PAN, LAN and WAN)	H.820–H.859
Multimedia e–health data exchange services	H.860–H.869

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

Recommendation ITU-T H.420

Telepresence system architecture

Summary

Recommendation ITU-T H.420 describes a high-level telepresence system architecture. It describes subsystem components: telepresence endpoints, multipoint control units, call and resource controllers, gateways, management systems and how these are used in a communication session to provide telepresence. It also provides guidance on telepresence room design in terms of room layout and room characteristics.

History

Edition	Recommendation	Approval	Study Group	Unique ID*
1.0	ITU-T H.420	2014-10-14	16	11.1002/1000/12245

* 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

Recommendation ITU-T H.420

Telepresence system architecture

1 Scope

This Recommendation first describes a telepresence system in a high-level technology agnostic manner taking into consideration the use cases and requirements defined in [ITU-T F.734]. This is then followed by a more in depth description of the components. The latter parts of this Recommendation then describe how [ITU-T H.323] relates to a telepresence system. It then details aspects regarding the telepresence room design.

Unless explicitly stated in the text, the interfaces depicted in the figures and functions in this Recommendation are logical and thus do not necessarily represent a particular physical implementation.

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 H.243] Recommendation ITU-T H.243 (2005), *Procedures for establishing communication between three or more audiovisual terminals using digital channels up to 1920 kbit/s.*

[ITU-T H.245] Recommendation ITU-T H.245 (2011), *Control protocol for multimedia communication.*

[ITU-T H.320] Recommendation ITU-T H.320 (2004), *Narrow-band visual telephone systems and terminal equipment*.

[ITU-T H.323] Recommendation ITU-T H.323 (2009), *Packet-based multimedia communications systems*.

[ISO 8995-1] ISO 8995-1 (2002), *Lighting of work places – Part 1: Indoor, including its Technical Corrigendum 1(2005)*.

3 Definitions

3.1 Terms defined elsewhere

This Recommendation uses the following terms defined elsewhere:

3.1.1 audio mixing [b-IETF RFC 7262]: Refers to the accumulation of scaled audio signals to produce a single audio stream. See [b-IETF RFC 5117].

3.1.2 layout [b-IETF RFC 7262]: How rendered media streams are spatially arranged with respect to each other on a telepresence endpoint with a single screen and a single loudspeaker, and how rendered media streams are arranged with respect to each other on a telepresence endpoint with multiple screens or loudspeakers.

NOTE – Audio as well as video is encompassed by the term layout – in other words, included is the placement of audio streams on speakers as well as video streams on video screens.

3.1.3 remote [b-IETF RFC 7262]: Sender and/or receiver on the other side of the communication channel (depending on context); i.e., not local. A remote can be an endpoint or an MCU.

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

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

3.2.1 endpoint: For the purposes of this Recommendation, an endpoint is a device representing the logical point of final termination through receiving, decoding and rendering and/or point of origin through capturing, encoding and sending of media streams, e.g., an ITU-T H.323 terminal.

3.2.2 telepresence media codec: This term (and function) is used to describe any media (i.e., audio or video) codec that would be applicable for use in a telepresence system.

3.2.3 telepresence system: 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 (MCUs), personal computers (PCs)) and environmental characteristics are used to establish telepresence.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

ACU Audio Capture Unit

ARU Audio Render Unit

CCA Chair-control Command Acquire

CCD Chair-control Command Disconnect

CCK Chair-control Command Kill

CCR Chair-control Command Release/Refuse

CRC Call and Resource Controller

GK Gatekeeper

GW Gateway

HDTV High Definition Television

HVAC Heating, Ventilation and Air Conditioning

IP Internet Protocol

ISDN Integrated Services Digital Network

IU Interactive Unit

LAN Local Area Network

LCD Liquid Crystal Display

MCU Multipoint Control Unit

MCV Multipoint Command Visualization-forcing

NAT Network Address Translation

NU	Network Unit
PC	Personal Computer
PDP	Plasma Display Panel
PLMN	Public Land Mobile Network
PSTN	Public Switched Telephone Network
QoE	Quality of Experience
QoS	Quality of Service
RAS	Registration, Admission and Status
RTCP	RTP Control Protocol
RTP	Real-Time Protocol
SIP	Session Initiation Protocol
TCU	Telepresence Control Unit
TMC	Telepresence Media Codec
TMIO	Telepresence Media Input/Output
TMP	Telepresence Media Processor
VCB	Video Command Broadcast
VCU	Video Capture Unit
VRU	Video Render Unit
WAN	Wide Area Network

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 takes place.

5.2 Icons used

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

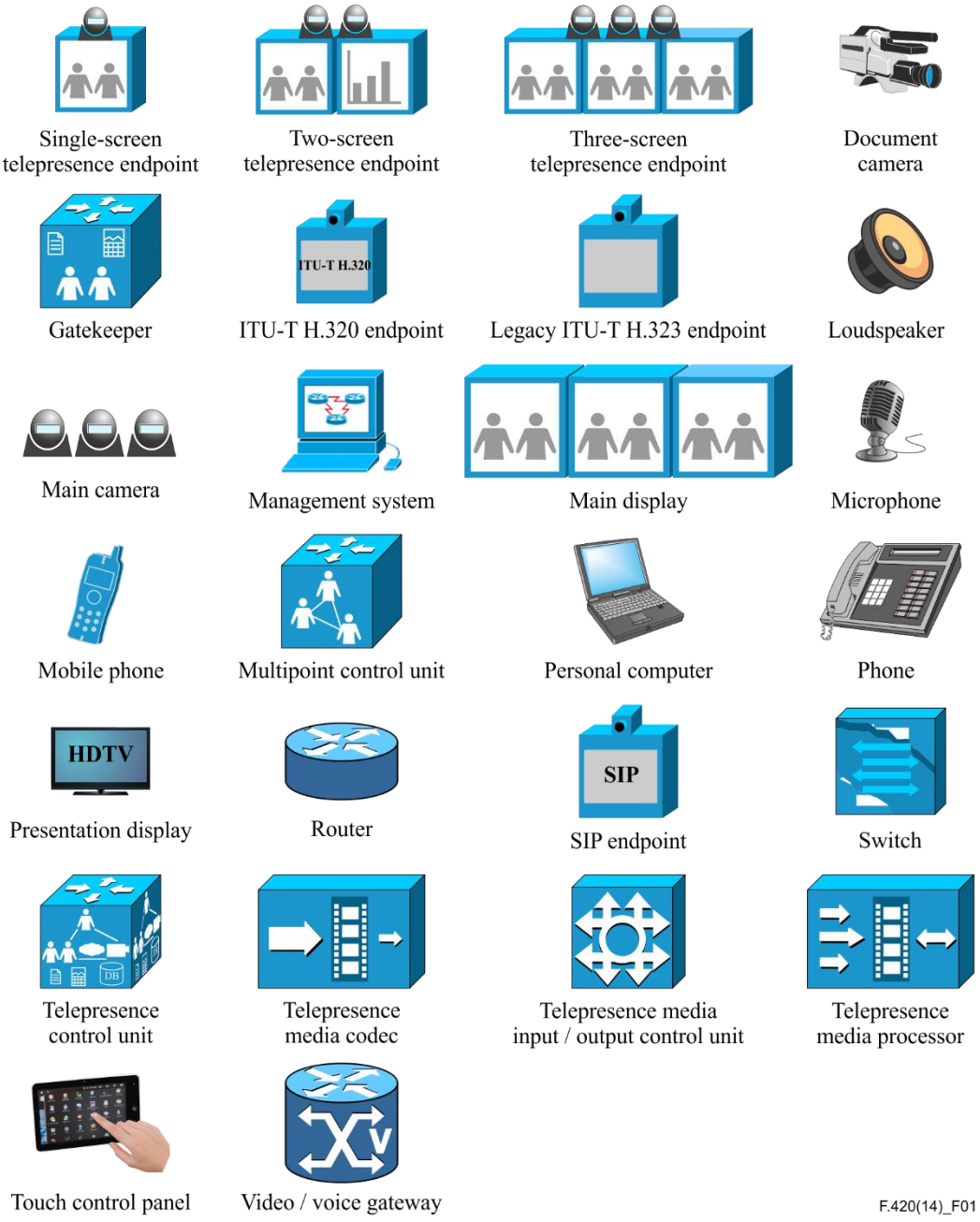

Figure 1 – Icons used in this Recommendation

6 High level architecture

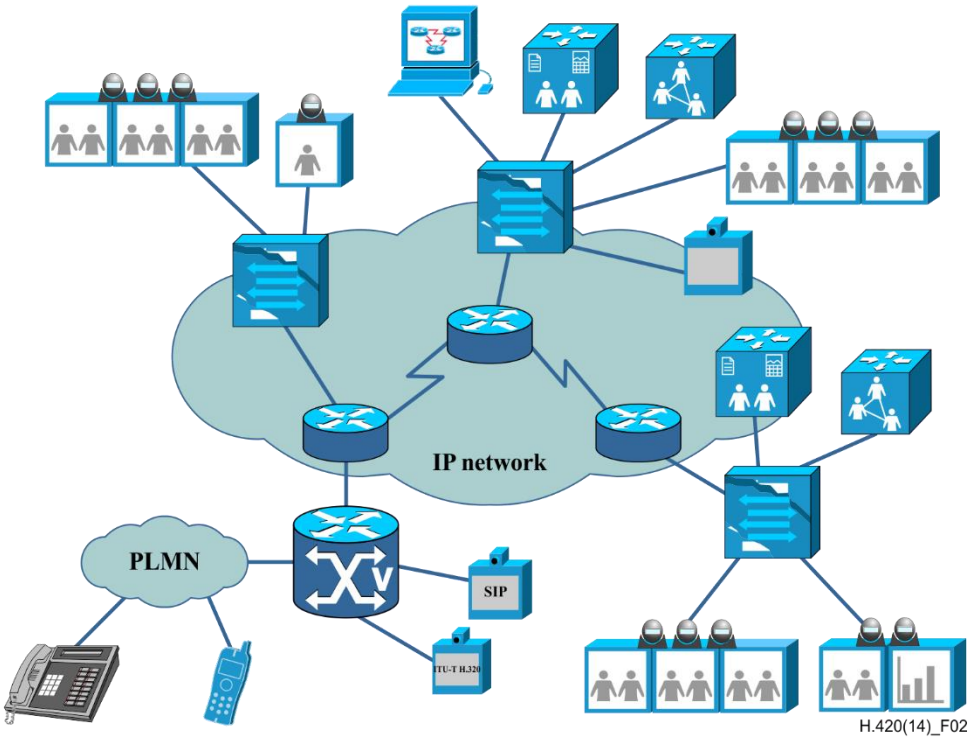

Figure 2 – High-level telepresence architecture network model

Figure 2 shows a high-level telepresence architecture including: telepresence endpoints, legacy video conferencing endpoints, legacy audio endpoints, legacy non-ITU-T H.323 endpoints including ITU-T H.320 and session initiation protocol (SIP) endpoints, multipoint control units (MCUs), call and resource controllers (CRCs), gateways (GWs), the management system, etc. There may be multiple forms of telepresence endpoints, such as three-screen endpoints and single-screen endpoints. The architecture supports inter-operation of heterogeneous telepresence endpoints and inter-operation between telepresence endpoints and legacy video conferencing and audio endpoints.

There may be multiple MCUs and CRCs in the architecture.

The architecture may have multiple GWs that connect to other types of networks, such as fixed and mobile telephone networks.

6.1 Description of subsystem components

A telepresence system is made up of several functional components. These components are illustrated in Figure 3.

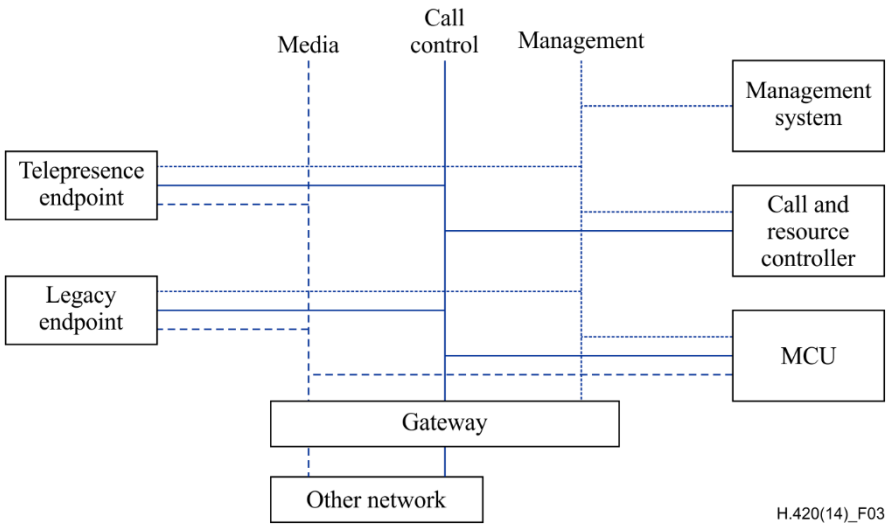

Figure 3 – Telepresence system functional components

Figure 3 shows telepresence system functional components and the call control and media interfaces between them. All interfaces are assumed to be bidirectional. There may be multiple instantiations of the components. In this case, there may also be additional call control or media interfaces between them. For example, there may be a call control interface between call and resource controllers. The exception is the management system. Whilst there may be multiple systems, this is treated as a monolithic system in this architecture.

The interfaces are numbered according to Tables 1 to 3.

The possible call control interfaces are illustrated in Table 1.

The possible management interfaces are illustrated in Table 2.

The possible media interfaces are illustrated in Table 3.

Table 1 – Telepresence system call control interfaces

Interface	Interface defined between	
	Component at one side	Component at the other side
C1	Telepresence endpoint	Call and resource controller
C2	Telepresence endpoint	MCU
C3	Telepresence endpoint	Gateway
C4	Telepresence endpoint	Legacy endpoint
C5	Telepresence endpoint	Telepresence endpoint
C6	Legacy endpoint	Call and resource controller
C7	Legacy endpoint	MCU
C8	Legacy endpoint	Gateway
C9	Legacy endpoint	Legacy endpoint
C10	Call and resource controller	MCU

Table 1 – Telepresence system call control interfaces

Interface	Interface defined between	
	Component at one side	**Component at the other side**
C11	Call and resource controller	Gateway
C12	Call and resource controller	Call and resource controller
C13	MCU	Gateway
C14	MCU	MCU

Table 2 – Telepresence system management interfaces

Interface	Interface defined between	
	Component at one side	**Component at the other side**
MT1	Call and resource controller	Management system
MT2	MCU	Management system
MT3	Gateway	Management system
MT4	Telepresence endpoint	Management system
MT5	Legacy endpoint	Management system

Table 3 – Telepresence system media interfaces

Interface	Interface defined between	
	Component at one side	**Component at the other side**
M1	Telepresence endpoint	MCU
M2	Telepresence endpoint	Gateway
M3	Telepresence endpoint	Legacy endpoint
M4	Telepresence endpoint	Telepresence endpoint
M5	Legacy endpoint	MCU
M6	Legacy endpoint	Gateway
M7	Legacy endpoint	Legacy endpoint
M8	MCU	Gateway
M9	MCU	MCU

6.1.1 Telepresence endpoint

The telepresence endpoint is the logical point of final termination, and may consist of more than one physical device. The main functions of the telepresence endpoint include:

– Setting up call sessions through capability negotiation,

– Capturing, encoding, packetizing and sending local media streams,

– Receiving, de-packetizing, decoding and rendering remote media streams,

– Managing ongoing sessions,

– Releasing sessions/resources.

In contrast to a telepresence endpoint, an MCU may also send and receive media streams, but it cannot initiate or terminate media streams through capturing and/or rendering. There may be multiple forms

of telepresence endpoints, such as: three-camera/screen endpoints, single-camera/screen endpoints, audio-only endpoints and etc.

Endpoint characteristics include:

– The number of main cameras, spatial relationships of cameras and capture capabilities, such as: resolution, frame rate and aspect ratio.

– Viewpoint and field of view for main cameras, including overlap regions, missing regions or whether images of cameras can be composited to a seamless image across the entire depth range. The number of presentation cameras and capture capabilities, such as: resolution, frame rate and aspect ratio.

– The video stream capabilities for data sharing such as: resolution, frame rate and aspect ratio.

– The number, type and size of main screens, spatial relationships of screens and screen capabilities, such as: resolution, frame rate, aspect ratio and screen border width.

– The number and type of presentation screens, spatial relationship of presentation screens and screen capabilities.

– The number of microphones and the spatial relationships between the captured audio signals from those microphones.

– The number of loudspeakers and the spatial relationships between the rendered audio signals.

– Display strategies for multipoint conference, such as the static or dynamic mapping of camera-to-display.

In general, endpoint characteristics are transmitted through signalling streams in the session initialization and used for capability negotiation to enable inter-operation of different telepresence endpoints. The characteristics of an endpoint may change during a session and may result in the need for additional media streams, modification of existing media streams or removal of the media streams.

When signalling its characteristics, an endpoint should take into consideration that it may be communicating with endpoints that have different capabilities and as such offer a range of characteristics to enhance interoperability between the endpoints. An endpoint should also offer characteristics that match the intended application. For example, if an application demands high resolution video, then an endpoint would not offer low resolution video.

NOTE – [ITU-T H.245] defines a terminal as "any endpoint and may be a user's terminal or some other communication system such as an MCU or an information server". As such, the term ITU-T H.323 telepresence terminal is equivalent to the term "telepresence endpoint".

6.1.2 Multipoint control unit

The multipoint control unit (MCU) is a device that connects two or more telepresence endpoints together to one single telepresence conference. An MCU can set up and control multiple videoconferencing call sessions. After receiving media streams from multiple endpoints or cascading MCUs and GWs, an MCU processes the media streams and sends the processed media stream to an endpoint, MCU or GW.

MCU must support video and audio mixing and media adaptation for different types of endpoints and should also support control functions, for example, auxiliary stream token control, user selection of video sources and composition, automatic video switching based on talker detection and other user and administrative control functions.

MCU should participate in the signalling negotiating the characteristics of the media in order to minimise the need for mixing or adapting media between endpoints where possible. This is to improve the QoS/QoE of the media.

6.1.3 Call and resource controller

A call and resource controller (CRC) is the resource management device for video conferencing and can provide resource management for endpoints, MCUs and GWs. The functions of a CRC include: address resolution, access control and bandwidth management. In ITU-T H.323 networks this may be realized as a gatekeeper and in SIP networks as a proxy server.

6.1.4 Gateway

A gateway connects various types of networks and provides interworking of signalling and media streams, such as interworking ITU-T H.323 and SIP. It may also provide adaptation for media streams that have different characteristics e.g., codec format. A gateway allows telepresence and non-telepresence endpoints and other components residing in different networks (e.g., intra and inter operator networks, operator and enterprise networks, etc.) to have access to the same conference.

6.1.5 Management system

A management system mainly contains control and management functions that provide specific telepresence services, such as device management, conference scheduling and user authentication, etc. It provides a statistic collection and reporting function for information relating to telepresence conference sessions and system usage.

6.1.6 Legacy conferencing endpoints

A legacy conferencing endpoint can be a legacy ITU-T H.323 or a non-ITU-T H.323 conferencing endpoint. In case of a legacy ITU-T H.323 endpoint, the main capabilities and functions are identical to the definition in [ITU-T H.323]. In case of a non-ITU-T H.323 endpoint, it can be a legacy endpoint which uses non-ITU-T H.323 open standards (i.e., SIP). These kinds of endpoints may be a SIP hard phone and SIP soft phone (i.e., PC or laptop computer). The legacy non-ITU-T H.323 endpoints can connect to other components (e.g., call and resource controller) through a gateway which performs protocol conversion and stream adaptation.

6.1.7 Network

A signalling and transport network underlies the telepresence system and connects the various network elements within or across the network boundaries (e.g., between private networks). The network is responsible for delivering signalling and media with an appropriate quality of service (QoS) in the telepresence system. The network elements via measurement of the network performance and/or through appropriate QoS mechanisms (e.g., packet prioritisation, resource reservation, media quality adaptation) should ensure an acceptable QoS for a telepresence experience. IPv4 or IPv6 network technologies may be applied to deliver this service.

The network may also be required to support mechanisms to secure telepresence session signalling and/or media flows. These mechanisms may be negotiated between network elements. These network elements may also require authentication in order to use the network resources.

7 Conference control functions

Conference control functions are those functions used to manage the display mode of screens and control cameras used in telepresence systems, support addition/deletion of participants in the telepresence session, control composition policy and switching of the meeting room media. The conference control functions described in this clause relate to the telepresence system as a whole. The functions may be contained in one or more telepresence subsystem components. For example, the removal of an endpoint from a conference may be initiated by an endpoint but needs an MCU to change its media mix to remove the media.

Conference control functions are grouped into three main groups:

– Media control: These functions relate to the negotiation and manipulation of the media streams.

– Meeting arrangement: These functions relate to the planning and running of a meeting.

– Endpoint management: These functions relate to the interaction of endpoints with a conference and the control of local functions.

The control functions described in these clauses are from a user perspective and provide an overview of call control functions available to the telepresence systems. A telepresence system may not support all conference control functions.

7.1 Media control functions

7.1.1 Chair control

The chair control functions are described in Table 4.

Table 4 – Chair control functions

Function	Description
Request chair control (makeMeChair)	Any endpoint can apply for chair control. Application is granted when there is no chair in the conference. If there is already a chair in the conference, granting of the application will depend on the decision of the current chair. NOTE – This is equivalent to the [ITU-T H.243] chair-control command Acquire (CCA).
Release chair control (cancelMakeMeChair)	Release chair control so that another endpoint/person may take the chair. NOTE – This is equivalent to the [ITU-T H.243] chair-control command release/refuse (CCR) command sent from the terminal.
Withdraw chair control (withdrawChairToken)	A previously chair controlled conference now has no chair control. NOTE – This is equivalent to the [ITU-T H.243] CCR command send from the MCU.
Floor request (requestForFloor)	Request the floor NOTE – This is equivalent to the [ITU-T H.243] chair-control command Kill (CCK).
Display floor requests (floorRequested)	Displays the list of endpoints/persons applying for the floor. NOTE – [ITU-T H.243] provides a facility to forward each request for the floor to the chair. It does not provide a facility to display the list of outstanding floor requests.

7.1.2 Media rendering

The media rendering functions are described in Table 5.

Table 5 – Media rendering functions

Function	Description
Select default view	This function selects the default view for an endpoint. This may be a certain endpoint, round-robin of endpoints when there are multiple endpoints, etc.
Select broadcasting endpoint (MakeTerminalBroadcaster)	All endpoints (except the origin) in the conference receive media from the selected broadcasting endpoint. NOTE – This is equivalent to the [ITU-T H.243] "Select Broadcaster" where the chair selects an endpoint/person's video to be distributed to the entire conference, i.e., command video command broadcast (VCB).
Cancel broadcasting endpoint (CancelMakeTerminalBroadcaster)	Media distribution from the broadcasting endpoint is cancelled and the conference reverts back to the previous mode of operation. NOTE – This is equivalent to the [ITU-T H.243] "Cancel Broadcaster" where the media distribution is cancelled and the conference reverts to a normal operating mode, i.e., command Cancel-VCB.
Select endpoint and captures	Only the selected media captures from a selected endpoint are distributed in the conference. Media capture from the other endpoints is stopped, i.e., microphones are muted.
Select a composed image	Manually set a composed image from more than one media capture.
Voice controlled switch	Indicate that the media to distribute in the conference is based on a volume threshold, e.g., loudest endpoint's media is distributed.
Automatic composition	Allow the system to compose an image from more than one media capture based on pre-set criteria.
Set presentation status	Indicates that an endpoint is to send presentation media.
Remove presentation status	Stop an endpoint from sending presentation media.
Request for visualization (BroadcastMyLogicalChannel)	An endpoint requests that its captures be distributed to the entire conference. NOTE – This is equivalent to the [ITU-T H.243] multipoint command visualization-forcing (MCV) command and is typically done when there is no chair.
Cancel visualization (CancelBroadcastMyLogicalChannel)	An endpoint cancels its previously granted visualization request. NOTE – This is equivalent to the [ITU-T H.243] Cancel-MCV command.

7.2 Meeting arrangement

The meeting arrangement functions are described in Table 6.

Table 6 – Meeting arrangement functions

Function	Description
Schedule conference	Add conference information, invite participants (Note 1).
Start conference	Start the conference instance. Enable calling of the endpoints. NOTE – For example, [ITU-T H.323] Call Setup message, conferenceGoal=invite.
Extend conference	Extend the end time of the conference (Note 1).
End conference	End the conference. All sessions with endpoints cease. NOTE – This is equivalent to the [ITU-T H.243] CCK command.
Lock conference	No more endpoints can join the conference (Note 2).
Unlock conference	Allow new endpoints to join a conference (Note 2).
NOTE 1 – There is no ITU-T Recommendation on conference scheduling. NOTE 2 – [ITU-T H.243] does not have this facility.	

7.3 Endpoint management

The endpoint management functions are described in Table 7.

Table 7 – Endpoint management functions

Function	Description
Call endpoint	Call one endpoint which is in the conference list but has not joined. Call all endpoints which are in the conference list but have not joined. NOTE – [ITU-T H.243] provides a conference list, but it only includes endpoints which are already in the conference.
Disconnect endpoint	End a session with an endpoint which has already joined in the conference. NOTE – This is equivalent to [ITU-T H.243] chair-control command disconnect (CCD) command.
Remove endpoint	Remove an endpoint from the conference list whether it has joined the conference or not. NOTE – [ITU-T H.243] provides a conference list and a facility for removing endpoints from the conference. But, the conference list only includes endpoints which are already in the conference.
Add endpoint	Add an endpoint to the conference list. NOTE – [ITU-T H.323] allows endpoints to be invited to an existing conference, but they are not added to the conference list until they are connected to the conference.
Mute	Mute and un-mute local or remote endpoint.
Set advertised captures	Set information associated with captures. Set which captures are advertised.
Far-end camera control	Move remote camera

8 Subsystem architectures

8.1 Endpoint architecture

The telepresence endpoint architecture is illustrated in Figure 4.

A telepresence endpoint consists of the following components:

– A video capture unit (VCU) that consists of multiple main cameras and presentation cameras (optional) in certain spatial relationships.

– A video render unit (VRU) that consists of multiple screens for main video streams and screens (optional) for presentation in certain spatial relationships.

– An audio capture unit (ACU) that consists of multiple microphones in certain spatial relationships.

– An audio render unit (ARU) that consist of multiple loudspeakers in certain spatial relationships.

– A telepresence control unit (TCU) that connects and controls other units. It includes: a telepresence media codec (TMC), telepresence media processor (TMP) and telepresence media input/output (TMIO) control units.

– A network unit (NU), including wired and wireless network interfaces and network switch, etc.

– An interactive unit (IU), including touch control panels, PCs, etc.

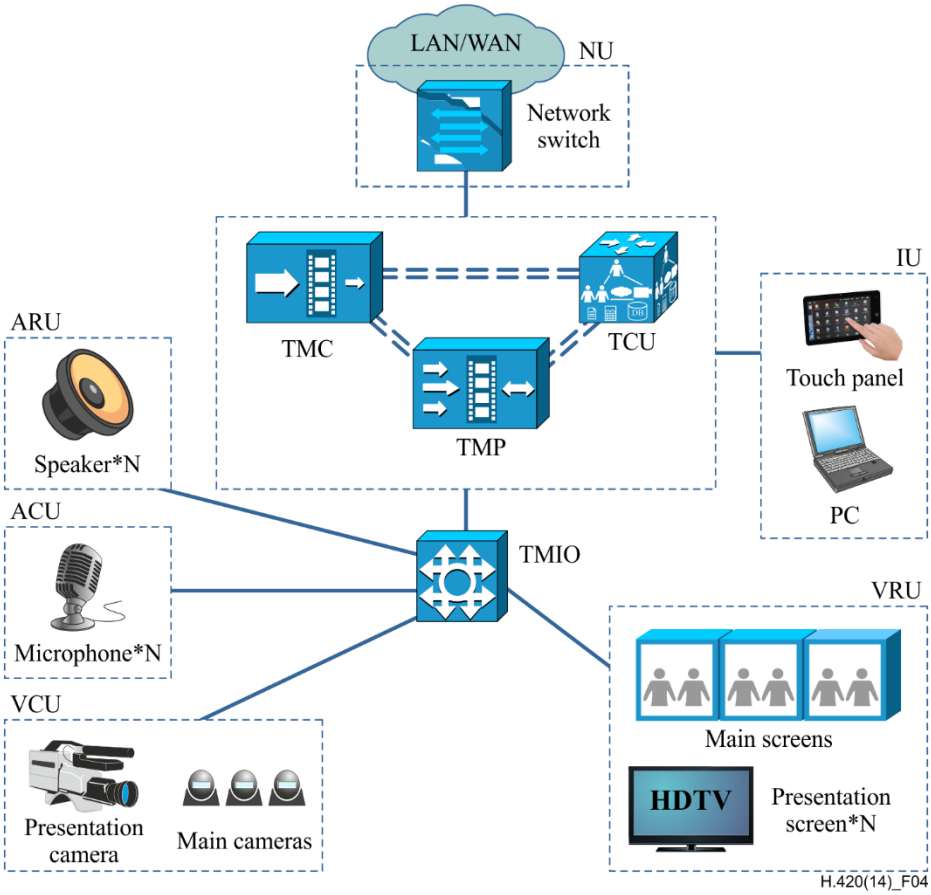

Figure 4 – Endpoint architecture

8.1.1 Endpoint architecture description

8.1.1.1 Video capture unit

A telepresence endpoint may have multiple video sources including main cameras, presentation cameras and auxiliary video stream(s) from a PC or other device. Video sources connect through a telepresence media input/output control unit that can select and control multiple video sources. See Figure 5.

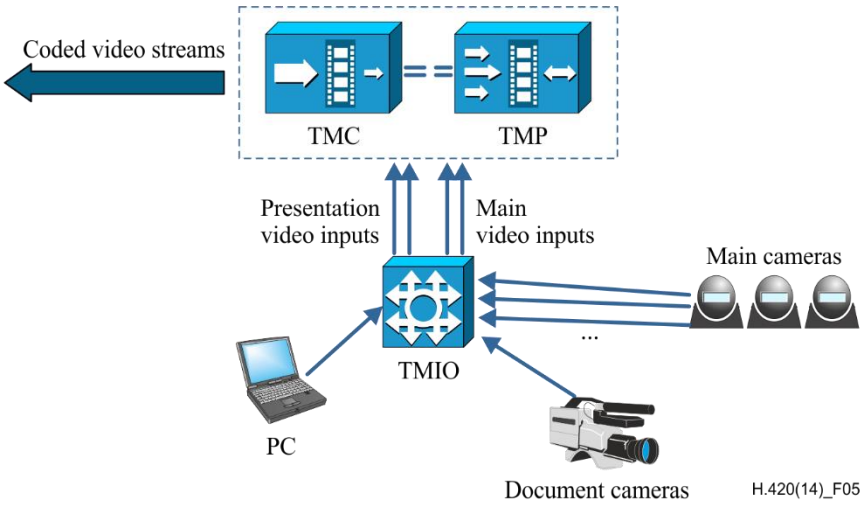

Figure 5 – Video capture unit architecture

Multiple main cameras may constitute a camera set for capturing images of participants at the telepresence endpoint. Each main camera captures images of participants in one region of the room corresponding to a particular spatial arrangement and the multiple images of cameras are combined to a wide view angle, high resolution view. Where spatial information is known, the video capture unit shall support the carriage of this information. A camera set may use one of the following three models: the converging camera model, the parallel camera model and the con-optical centre camera model. See Figure 6 for an illustration of various camera models.

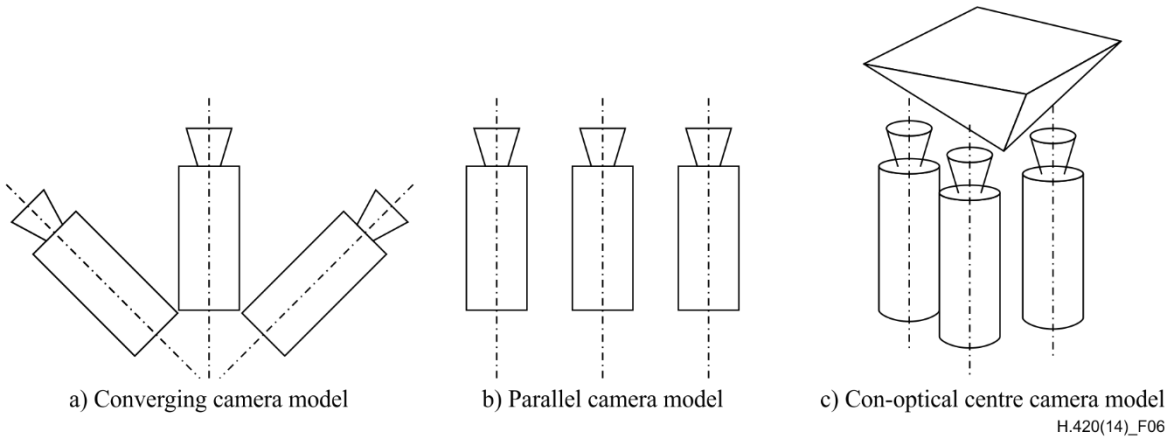

Figure 6 – Camera models

In the converging camera model, each of the multiple cameras looks out directly in a different direction. For the parallel camera model, in contrast to the converging camera model, all optical camera axes have the same direction. Multiple cameras capture multiple virtual images reflected by a mirror and the captured images almost have the same optical centre in the con-optical centre camera

model. For the first two camera models (i.e., the converging camera model and the parallel camera model), there may be overlap and missing regions between two images and individual images may not be able to be integrated seamlessly unless objects are extremely distant. While in the con-optical centre camera model, individual images can be stitched to a seamless image over the entire depth range. The composited image can be scaled to an image having a single camera's resolution for inter-operation with a legacy videoconferencing endpoint that does not support this capability.

Presentation cameras are used to capture images of documents or objects in a conference room. There may be various spatial placements for presentation cameras, for example, hanging the cameras from the ceiling to capture images of objects on a desk.

Auxiliary video streams from a PC or other devices are used for sharing data.

8.1.1.2 Video render unit

A video render unit may comprise of one or more main screens and presentation screens, see Figure 7. Main screens are primarily used to display the video streams captured by local and remote main cameras and can be liquid crystal display (LCD), plasma display panel (PDP) monitors or projectors. In a fully immersive system, a main screen should have sufficient size to display actual-sized images. Multiple monitors are usually integrated to present a panoramic view, or a single projection screen and multiple projectors can also be used.

The telepresence media input/output control unit can dynamically arrange display patterns and contents for main screens and presentation screens in an automatic or manual way:

– Main screens may display the local and remote main video streams and presentation video streams with reproduction in a spatially correct manner based on the spatial information associated with the streams.

– Typically where a presentation screen does not have a camera capable of capturing the participant(s) it should only display the local and remote presentation video streams, because eye contact cannot be achieved when a main video stream is presented on a presentation screen.

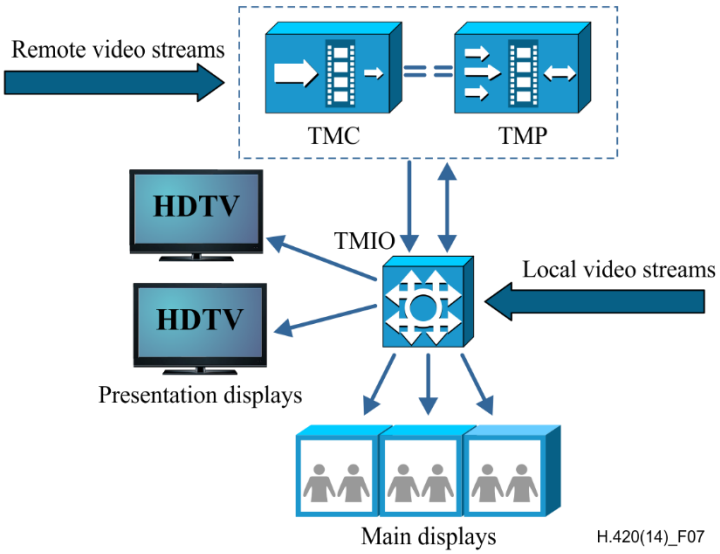

Figure 7 – Video render unit architecture

8.1.1.3 Audio capture unit

An audio capture unit may comprise of multiple microphones in some spatial placement for capturing audio signals at the telepresence endpoint, see Figure 8. This spatial placement would largely be dependent on the telepresence endpoint configuration. For example, microphone placement may differ between single-row and multiple-row configurations. There may be no direct mapping between

microphones and audio streams. Where spatial information is required, the audio capture unit should support the carriage of this information.

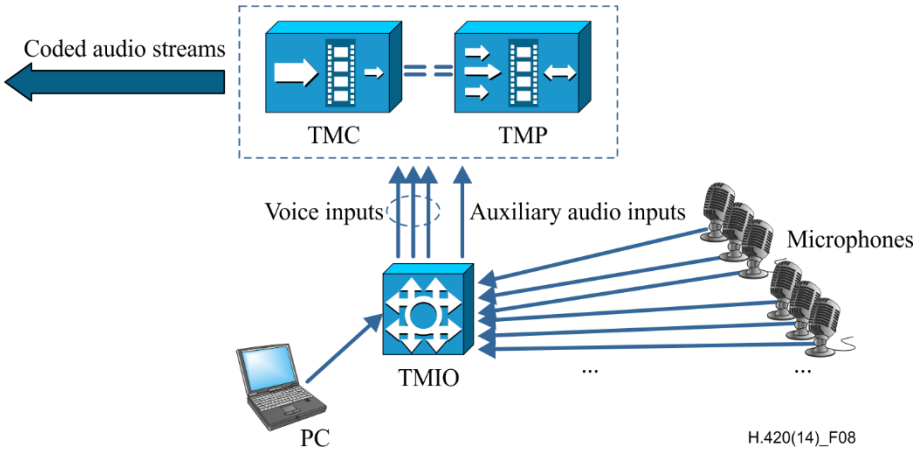

Figure 8 – Audio capture unit architecture

Microphones may connect through a media input/output control unit that can select and control multiple audio sources. Auxiliary audio streams from a PC or other devices also connect to an endpoint through the media input/output control unit. Audio signals from multiple microphones may be combined into one single audio stream for transmission. For example, the audio signal from each microphone may be coded as a separate channel using an applicable audio codec. These audio encodings are then transported as an aggregated stream to the remote endpoint via a suitable transport such as real-time protocol (RTP).

8.1.1.4 Audio render unit

An audio render unit may comprise of multiple loudspeakers in some spatial placement for rendering audio signals of the remote telepresence endpoint, see Figure 9. There are no direct mappings between loudspeakers, microphones and screens, but the spatial information of sound from the remote endpoint can be reproduced through the audio render unit. Both the sound and the image positions of remote participants should coincide.

Loudspeakers should connect through the media input/output control unit that can select and control audio output dynamically.

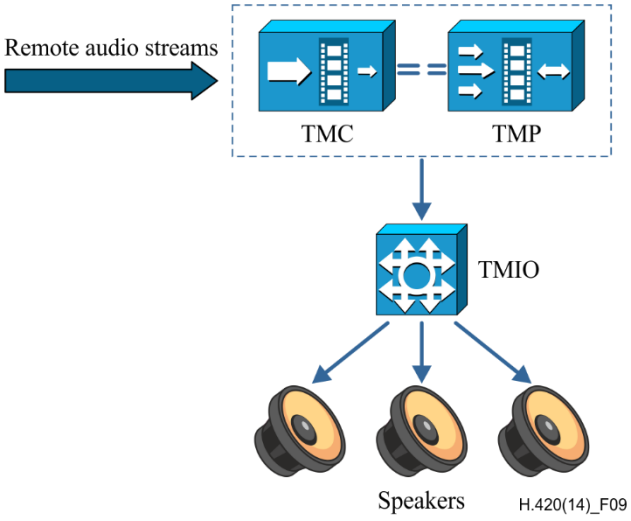

Figure 9 – Audio render unit architecture

8.1.1.5 Telepresence control unit

A telepresence control unit connects other units in a telepresence endpoint. The main functions of the telepresence control unit include:

– To provide interfaces to other functional units in an endpoint, invoking other units to complete processing.

– To handle signalling processes, including: registration, authentication, session initialization, session control and session termination, capabilities negotiation for inter-operation, etc.

– To send and receive multiple media streams.

– To provide contacts and other interfaces.

– To provide device management and debug interfaces.

8.1.1.6 Telepresence media codec

The telepresence media codec is a logical unit that is used for coding and decoding of multiple video and audio streams. Synchronization between video and audio streams is supported. The telepresence media codec can be implemented using two configurations; a centralized solution or distributed solution. A centralized solution uses a single device to code and decode multiple media streams. Synchronizing media streams for this configuration is simpler but the device requires higher processing capabilities. The distributed solution needs more than one device for coding and decoding of multiple media streams which results in more complicated synchronization strategies. However, this configuration may use devices that have lower processing capabilities than that of the centralized solution.

8.1.1.7 Telepresence media input/output control unit

The telepresence media input/output control unit is a logical unit that can select and control multiple input/output media streams that include main and auxiliary video, audio and text streams. Video and audio capture units, video and audio render units, PCs, text terminals and other audio endpoints are connected to the telepresence media input/output control unit.

8.1.1.8 Telepresence media processor

The telepresence media processor is used for processing video and audio steams to provide the required audio and visual experience. The main functions of telepresence media processor include:

– Geometric correction for multiple camera images to get better image mosaic effect.

– Colour correction for making luminance and chrominance of multiple camera images uniform.

– Geometric correction for multiple video stream rendering, which adapts video outputs to various types of screens to produce a better display performance, for example, geometric correction for a curved projection screen.

– Colour correction for multiple video stream rendering for a better rendering performance.

– Pre-processing for audio signals captured by audio capture unit, such as audio de-noising and enhancement for multiple microphones, etc.

– Pre-processing for audio signals rendered by an audio render unit, for example, improving audio quality of loudspeakers.

– Incorporation of text information (e.g., captioning) into video streams prior to encoding or post decoding.

8.1.1.9 Interactive Unit

The interactive unit provides an interface for users to receive information regarding a telepresence conference and to be able to control various functions:

– call control: For example, establishing and releasing a telepresence conference,

– media control: For example, selecting the media format and changing rendering properties such as display position,

– conference control: For example, managing the meeting via chair control,

– collaboration: For example, setting up and managing white boarding.

The interactive unit may be realized through one or more physical devices, for example, touch control panels, tablets, etc. The goal of the interactive unit interface is to allow participants to control a telepresence conference with the minimum of input.

8.2 Multipoint control unit

The main functions of an MCU are basically identical to the definitions in [ITU-T H.323] as per the following descriptions.

Access functions

– To support accesses of various types of endpoints that include: multi-screen telepresence endpoint, single-screen telepresence endpoint and audio endpoints.

– To support multi-image composition videoconferencing.

– To support single/multiple auxiliary media streams.

– To support MCU cascading connection. Multiple media channels should be supported and can be flexibly multiplexed in cascading connection mode.

– To assure QoS.

– To support automatic network address translation (NAT) traversal

– To support both IPv4 and IPv6 connections.

Adaptation functions

– Adaptation for media streams according to the capabilities of the endpoints, such as video transcoding, audio transcoding, etc.

– Adaptation for media streams according to the endpoint characteristics, such as screen aspect ratio, etc.

– Utilising endpoint characteristics to determine cases where no adaptation of the media stream is necessary.

Control functions

– To support voice switch and manual switch mode.

– To support switch based on the endpoint or the screens when an endpoint has multiple screens.

– To support conference chair control, such as media selection, adding or deleting endpoints.

– To support far-end camera control.

– To support non-telepresence heterogeneous systems.

– To support spatial arrangement.

Security functions

– To support authentication and signalling/media security.

8.3 Call and resource controller

The call and resource controller (CRC) is an optional function that provides call and resource control functions including (but not limited to):

– Address translation
– Admission control
– Bandwidth control
– Zone management
– Call authentication and authorisation
– Call management/routing
– Digit translation
– Call feature provision

Multiple call and resource controllers may be utilized in a telepresence system.

Examples of CRCs are ITU-T H.323 gatekeepers and SIP proxies.

8.4 Gateway

The main gateway functions are identical to those defined in [ITU-T H.323].

9 ITU-T H.323 architecture

9.1 Mapping telepresence to ITU-T H.323 architecture

The telepresence system architecture is based on the existing [ITU-T H.323] system. A telepresence endpoint can be viewed as another kind of ITU-T H.323 endpoint, as depicted in the Figure 10.

Figure 11 illustrates the interfaces between the different components of a multimedia system based on [ITU-T H.323].

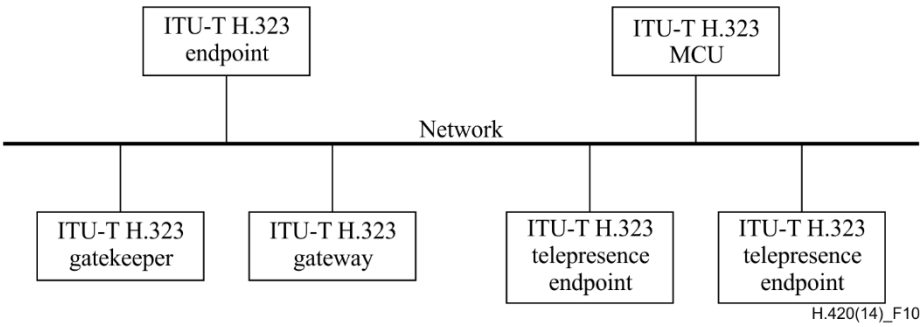

Figure 10 – ITU-T H.323 telepresence network model

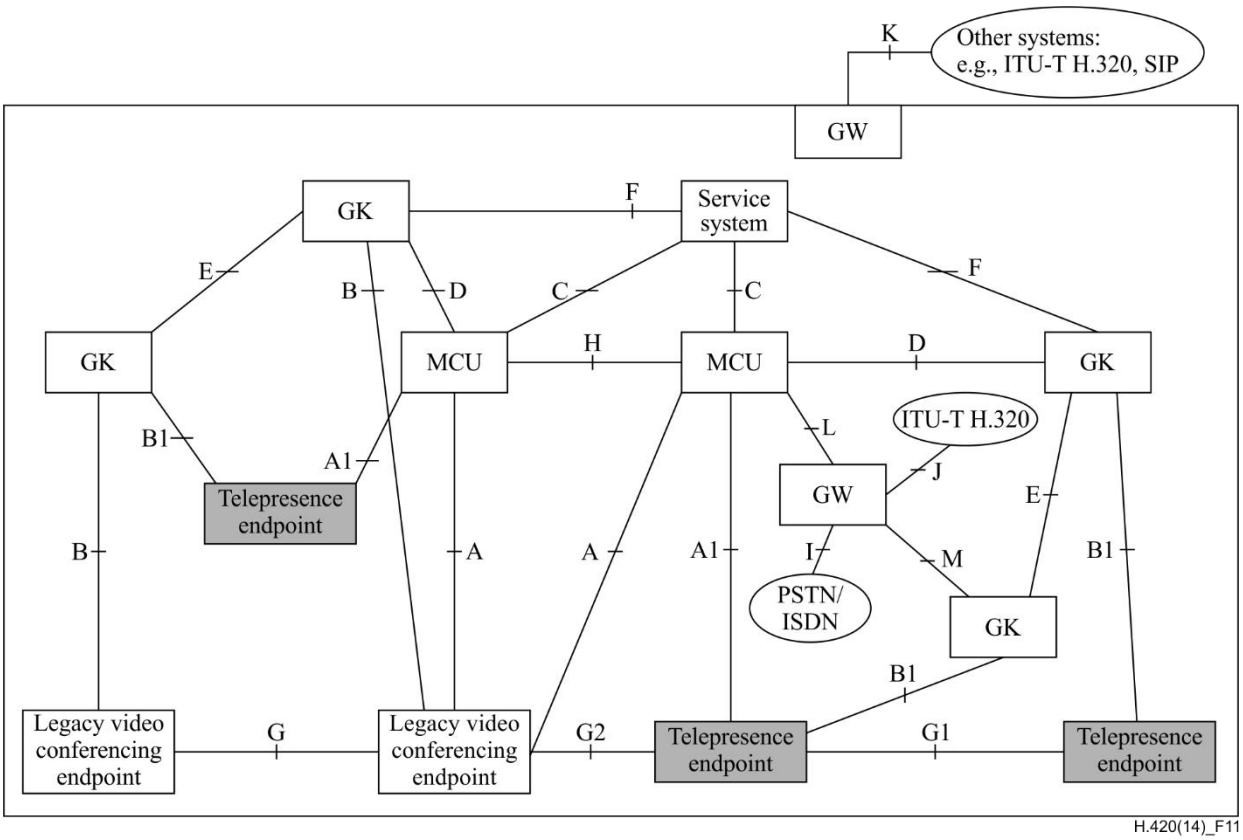

Figure 11 – ITU-T H.323 telepresence system architecture

The main interfaces (reference points) in ITU-T H.323 telepresence system architecture are:

– A: Interface between the legacy video conferencing endpoint and the MCU for transmitting conference call and control information, completing creation and release of media channels and media interaction. This relates to the generic telepresence system interface C7 (Table 1).

– B: Interface between the legacy video conferencing endpoint and the gatekeeper (GK) for transmitting registration, authentication and call process information using registration, admission and status (RAS) messages. This relates to the generic telepresence system interface C6 (Table 1).

– C: Interface between the MCU and the service system for transmitting conference scheduling information, creating and ending conferences. This relates to the generic telepresence system interface MT2 (Table 2).

– D: Interface between the MCU and the GK for registration, authentication, call process of MCU and transmitting resolving information of user addresses. This relates to the generic telepresence system interface C10 (Table 1).

– E: Interface between the GKs, transmitting resolving information of user addresses using RAS messages, etc. This relates to the generic telepresence system interface C12 (Table 1).

– F: Interface between the GK and the service system for transmitting GK's management and configuration information, user information, billing data, etc. This relates to the generic telepresence system interface MT1 (Table 2).

– G: Interface between the legacy videoconferencing endpoints for transmitting call signalling, media streams and related control information when a terminal has a peer-to-peer communication. This relates to the generic telepresence system interfaces C9 (Table 1) and M7 (Table 3).

– H: Interface between the MCUs for transmitting media streams between MCUs and information of the MCU cascade connection and inter-operation. This relates to the generic telepresence system interfaces C14 (Table 1) and M9 (Table 3).
– I: Interface between the gateway (GW) and the PSTN/ISDN network.
– J: Interface between the GW and the [ITU-T H.320] network for inter-operation between [ITU-T H.323] and [ITU-T H.320] service network.
– K: Interface with other systems. This relates to the generic telepresence system interfaces C13 (Table 1) and M8 (Table 3).
– L: Interface between the GW and the MCU. This relates to the generic telepresence system interfaces C13 (Table 1) and M8 (Table 3).
– M: Interface between the GW and the GK. This relates to the generic telepresence system interfaces C11 (Table 1).

In order to support telepresence endpoints, several new interfaces based on the existing ones are introduced:

– A1: Interface between the telepresence endpoint and the MCU for transmitting call and control information of telepresence conference, completing creation and release of media channels and exchanging video and audio streams between telepresence system and MCU. This relates to the generic telepresence system interfaces C2 (Table 1) and M1 (Table 3).
– B1: Interface between the telepresence endpoint and the GK for transmitting registration, authentication and call process information using RAS messages. This relates to the generic telepresence system interface C1 (Table 1).
– G1: Interface between the telepresence endpoints for transmitting call signalling, media streams and related control information when a telepresence endpoint has a peer-to-peer communication with another one. This relates to the generic telepresence system interfaces C5 (Table 1) and M4 (Table 3).
– G2: Interface between the legacy video conference endpoint and the telepresence endpoint for transmitting call signalling, media streams and related control information when a telepresence endpoint has a peer-to-peer communication with a legacy video conference endpoint. This relates to the generic telepresence system interfaces C4 (Table 1) and M3 (Table 3).

9.1.1 Peer-to-peer communication

Figure 12 illustrates a peer-to-peer communication scenario for telepresence.

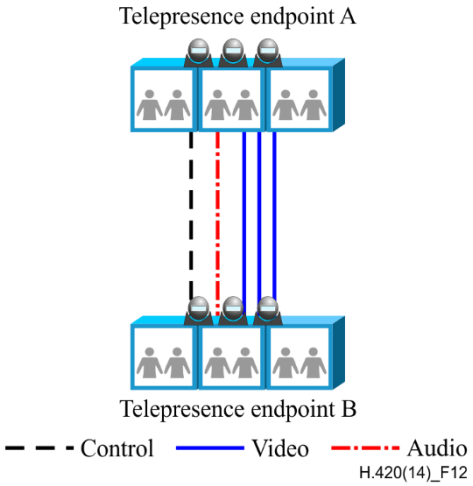

Figure 12 – ITU-T H.323 telepresence peer-to-peer communication

The communication between telepresence endpoints will have a single call control signalling channel supplemented by additional control signalling, one or more video streams and one or more audio streams.

The control signalling stream is used for controlling the whole calling process, for example, enabling/disabling the sending of auxiliary streams.

9.1.2 Multipoint communication

Figure 13 illustrates a multipoint communication scenario for telepresence.

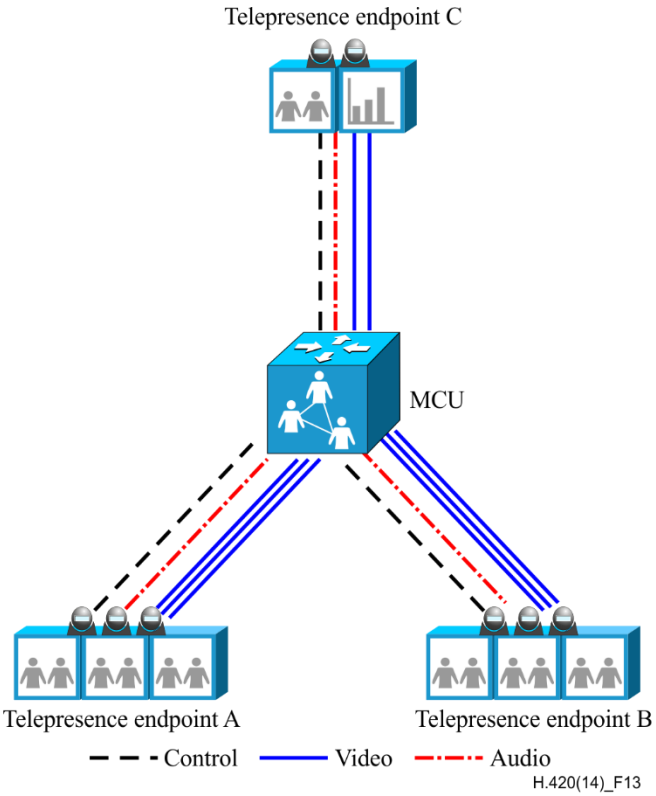

Figure 13 – ITU-T H.323 telepresence multipoint communication

Different types of telepresence endpoints may connect to an MCU.

Telepresence endpoints send media streams according to the results of capability negotiation.

The MCU completes the adaptations for media streams, for example, adaptation for single channel and multichannel audio streams, retransmission for multiple video streams and adaptation for conference control.

9.2 Call process

The call process of an ITU-T H.323 based telepresence system follows the basic call phases as outlined in clause 8 of [ITU-T H.323]. However, there are additional functionalities in the call phases in order to support telepresence. The additions are outlined in Table 8 and Figure 14 illustrates the call phases.

Table 8 – Additional functionalities in call phases in order to support telepresence

Phase A:	An ability to negotiate the specific features related to the use of telepresence information is added. This allows an endpoint, gateway or MCU to indicate whether or not it supports telepresence. This is beneficial in order to determine early on in the call process whether a peer supports telepresence and allows endpoints to optimise their signalling, i.e., the endpoint would not send telepresence information if it is not used.
Phase B:	An ability to link the ITU-T H.245 terminalCapabilities (i.e., media characteristics) with a set of telepresence information is added. This allows an endpoint, gateway or MCU to link a set of terminalCapabilities with a set of telepresence information. For example, it enables the characteristics of the capture media (e.g., spatial information) to be linked to a particular potential media stream. There is also an ability to transfer and negotiate the telepresence information (e.g., information regarding the media captures).
Phase C:	An ability to indicate how the ITU-T H.245 logical channels (and thus the media) relate to a particular set of telepresence information is added.
Phase D:	The indication of the support of telepresence and the support telepresence information may be required for particular services, e.g., multipoint conferencing scenarios.
Phase E:	An ability to remove telepresence information from a media stream may be required in addition to normal release procedures.

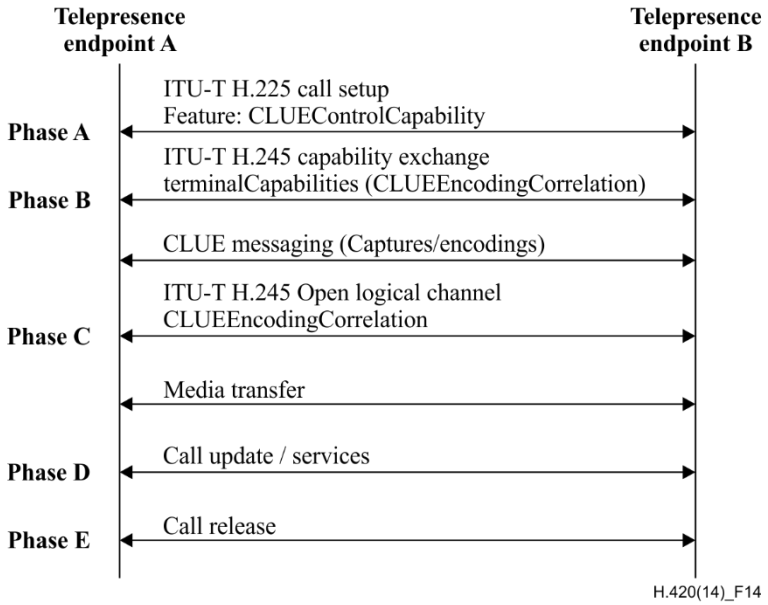

Figure 14 – ITU-T H.323 telepresence call processes

10 Telepresence room design

10.1 Room layouts

This clause details several example room layouts capturing and rendering room participants in a manner that facilitates a telepresence experience. Other configurations are possible.

In these examples where the multiple capture regions are utilized, the boundaries between the regions may be convergent or appear to be parallel. These boundaries may have an impact on the apparent gap between renderings of the captures.

10.1.1 Single row seating three-screen telepresence system

Figure 15 depicts the room layout of a single row three-screen telepresence system, with the following characteristics:

– Three main cameras and three main screens.

– Three regions for participants in the room, and each region contains two seats.

– Each main camera captures images of one region.

The missing/overlapped region of two captured images is located at the corner of the conference table and covered by borders of the two main screens.

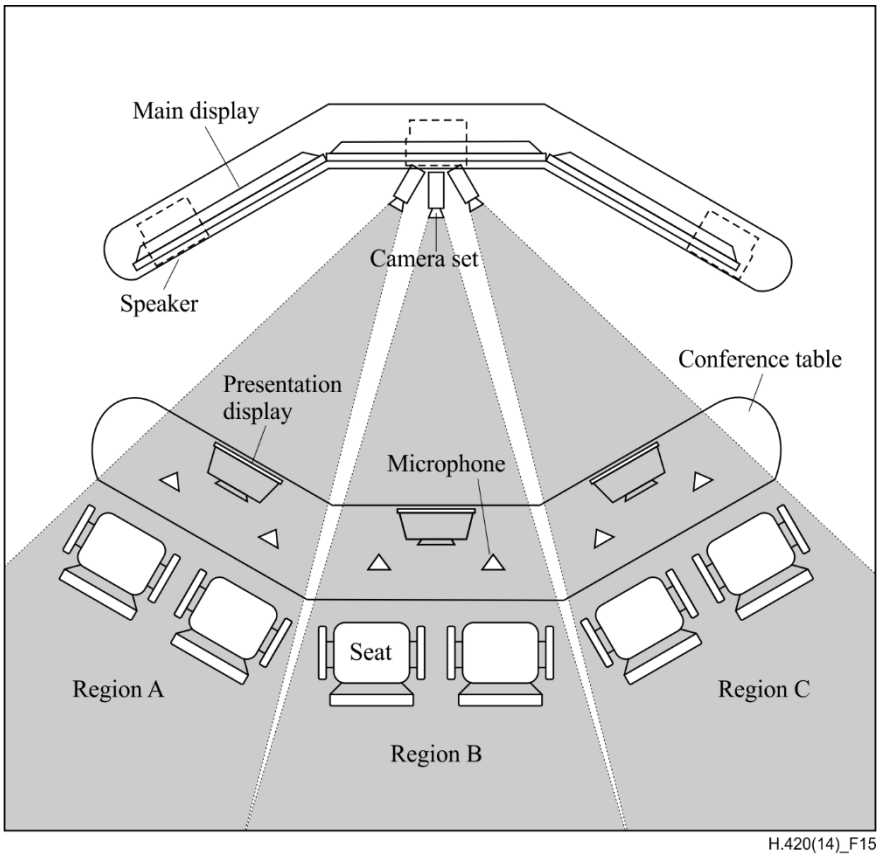

Figure 15 – Three-screen telepresence room layout (single row)

10.1.2 Multi-row seating three-screen telepresence system

Figure 16 depicts the room layout of a multi-row three-screen telepresence system, with the following characteristics:

– Three main cameras and three main screens.

– Three regions for participants in the room, and each region contains five seats.

– Each main camera captures images of one region.

The missing/overlapped region of the two captured images is located at the corner of the conference table and is matched to the borders between the two main screens.

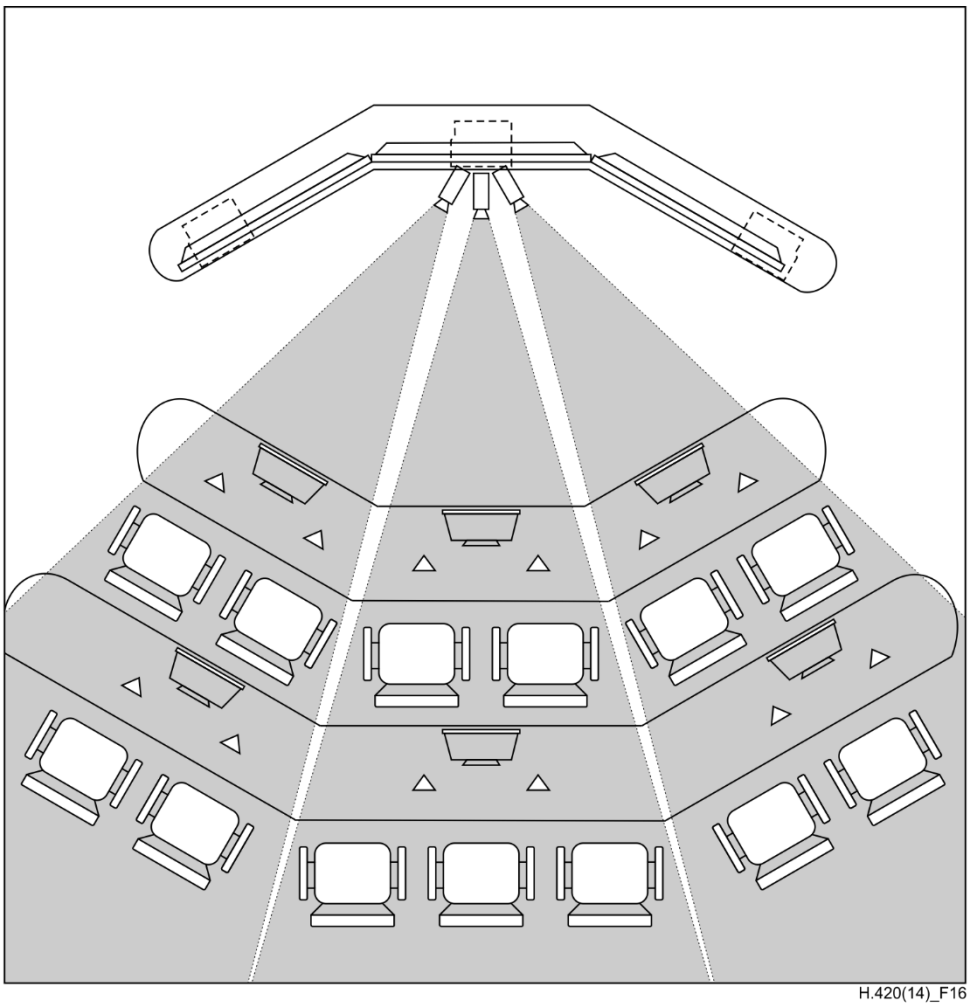

Figure 16 – Three-screen telepresence room layout (multi-row)

10.1.3 Multi-row seating four-screen telepresence system

Figure 17 depicts the room layout of a multi-row four-screen telepresence system, with the following characteristics:

– Four main cameras and four projectors.

– Four regions for participants in the room, and each region contains four or five seats.

– Each main camera captures images of one region.

– Three overhead microphones capturing multichannel audio from the room.

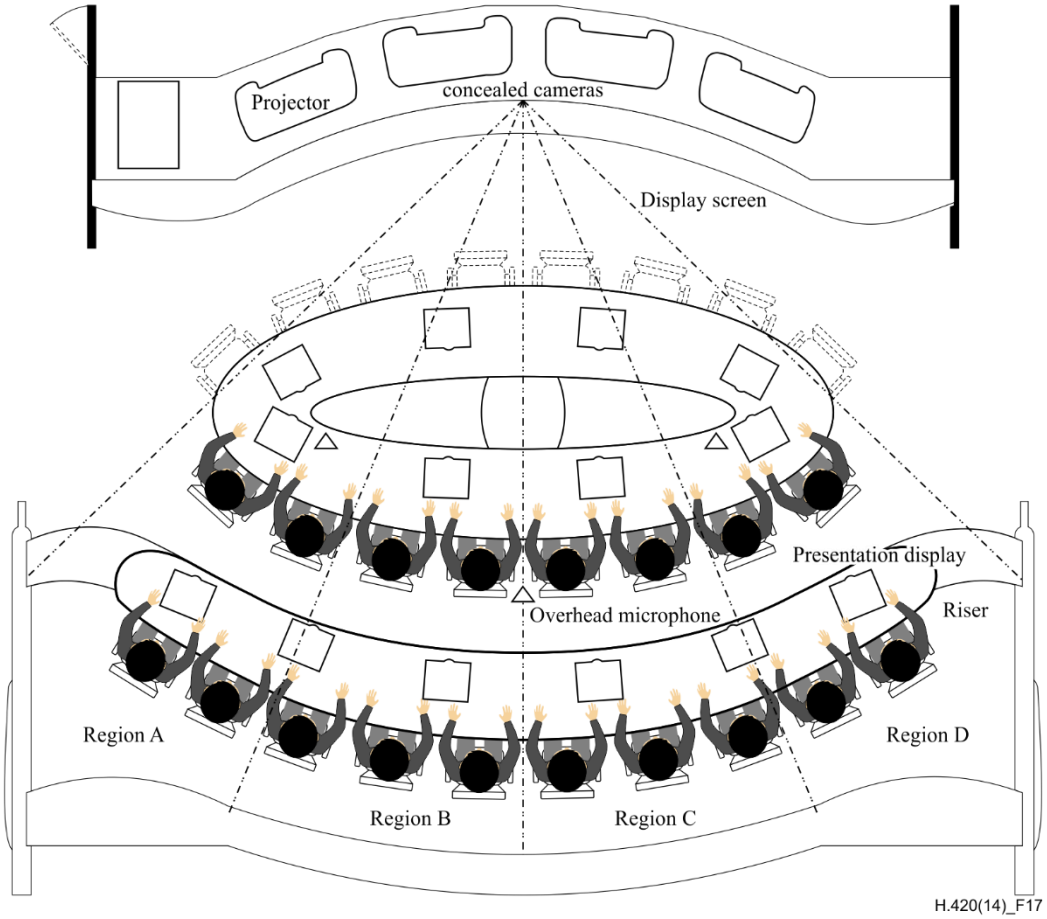

Figure 17 – Four-screen telepresence room layout (multi-row)

Each projected image shows the far-end participants at their true size. The display height is chosen so that standing individuals of normal height are displayed without cropping. In order to increase the height, 4:3 aspect ratios are used. Since projectors are used, there is no gap between the displayed images.

Main cameras are placed behind the projector screens, and are placed at eye level. The cameras are aligned to minimise gaps between captures. The seating is organised so that a person does not span a gap between captures.

The second row of participants is elevated about 25 cm above the first row. This allows all participants to be easily seen.

10.1.4 Single row seating six-screen telepresence system

Figures 18 and 19 depict the room layout of a six-screen telepresence system, with the following characteristics:

– Three main cameras and six screens with narrow frame borders.

– Three screens in a row, two rows of screens.

– Main cameras placed between the screens of the bottom and top row.

– Three regions for participants in the room, and each region containing two or three seats.

– Each main camera captures images of one region.

– One wall projector displays a selected presentation/stream.

As can be seen in Figure 18, a set of three screens are arranged in one row. Each screen shows the far-end participants at their true size. Three main cameras are placed at eye level and sandwiched

between two rows of screens; each camera is placed in the centre of the bottom frame of the higher screen and in the centre of the top frame of the lower screen.

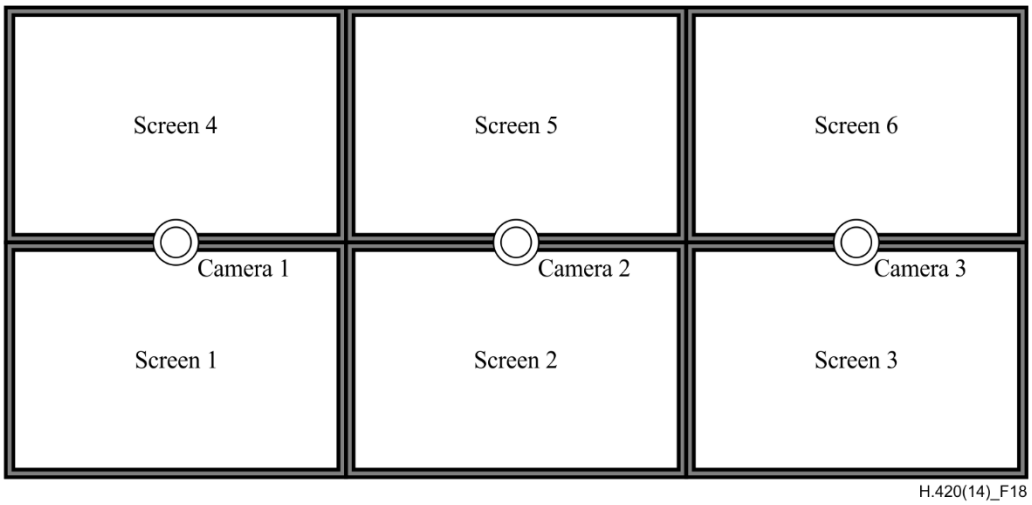

Figure 18 – Screen and camera layout for six-screen telepresence

The use of two rows of screens allows for other display scenarios not possible with a single row of screens. This is particularly useful in scenarios with more than two end points in a call. For example, the following scenarios are possible:

a) The simultaneous display of two endpoints, with one endpoint shown in the first row and the other in the second row;

b) Four adjacent screens, e.g., the left two and the middle two, displaying a scaled image of the current speaker, and the remaining two screens displaying in life size;

c) All the screens displaying switched speakers, behaving like a switched video wall.

NOTE 1 – This is possible in a single row. An additional row allows for more participants.

d) The upper row displaying active speakers while the lower row displays presentation and auxiliary information;

e) The upper row displaying the local endpoint while the lower row displays the remote endpoint.

As shown in Figure 19, each main camera captures a region of two (or three) seats along a table. Presentation screens and microphones are mounted on the long table. An additional wall projector is provided to display desired streams/presentations upon user demand.

Since screens with narrow frame borders are used, the displayed images need more manipulations to obtain a continuous view. So image processing methods, such as cropping and scaling, are required to provide a better visual experience when interoperating with screens with broader frame borders.

NOTE 2 – The dashed lines in Figure 19 represents objects that may be used when operating in a non-telepresence mode.

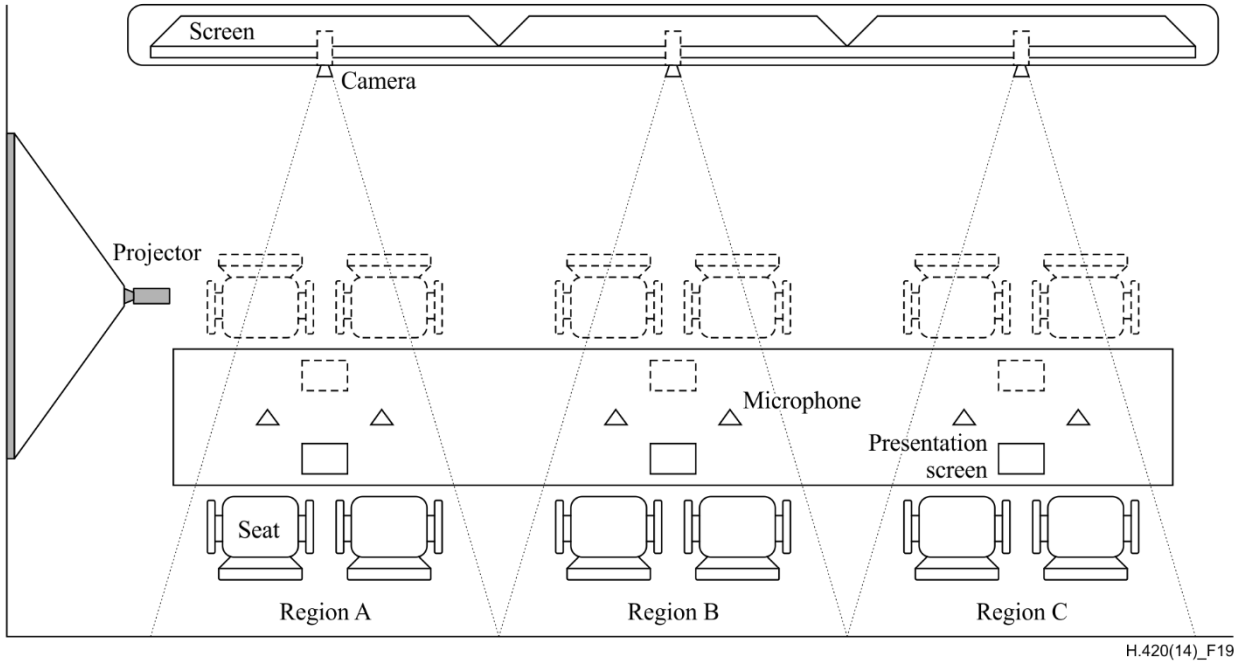

Figure 19 – Six-screen telepresence room layout

10.2 Room characteristics

This clause discusses room design considerations when the room is used as part of a telepresence system. Many of the considerations apply for non-telepresence conferencing systems.

10.2.1 Room dimensions and shape

There are several factors that dictate the dimensions of a room associated with a telepresence system. The size of the room (width, depth and height) should be able to accommodate:

– Telepresence system hardware.
– Furniture associated with the telepresence participants, e.g., desks, chairs.
– Furniture associated with other participants in the room, e.g., chairs, couches.
– Ancillary furniture associated with the room, e.g., tables, cupboards.
– Extra devices such as additional monitors/document camera, etc.
– Circulation areas for the participants to access the telepresence system and furniture.
– Allowance for access to the telepresence hardware or other services for maintenance.
– Allowance for fittings and services, e.g., lighting, heating, ventilation and air conditioning (HVAC) controls, fire safety equipment.

The placement of the telepresence participants should take into consideration the design of the capture devices. Cameras have a certain depth of field and capture area in both the horizontal and vertical planes. Therefore, there is a minimum and a maximum distance where a participant can be located. Likewise, microphones have a certain capture area.

The layout (horizontal and vertical) and the position of participants in relation to cameras and displays should be considered so that both direct and indirect gaze awareness is maximized.

It should also be recognized that rooms that are in excess of the size needed for the telepresence system with the above considerations present some problems. For example, a telepresence system may be accommodated in an auditorium. In this case, even and consistent lighting can be a challenge. A larger room may also present acoustic problems.

The shape of the room also needs to be considered. Doors, windows, the building fabric (such as columns) will dictate where the telepresence system hardware can be placed. Asymmetric or curved walls and ceilings will have acoustic impacts as well as presenting issues with capture video. Doors and windows present acoustic issues in allowing outside sound into a room and reverberation. In addition windows present lighting issues.

Telepresence aims to provide a being there experience. Therefore, the above needs to be considered so that a telepresence system can provide a pleasing experience for the participants.

10.2.2 Wall, floor and ceiling surfaces

The wall material, texture and colour influence both audio and video aspects of a telepresence system.

The type of wall material influences the acoustic properties of the room. Differing materials reflect and transmit sound at different rates. Materials should be chosen to maximise the amount of sound being absorbed. Where the wall material cannot be chosen, it may be possible to apply a surface treatment to achieve the desired characteristics.

The wall texture and any patterning may influence video encoding. Walls with a "busy" pattern or texture may cause problems for video coding. Plain surfaces should be used.

The wall colour also affects the room lighting in terms of the amount of light reflected and the colour of the reflected light. In order to provide a being there experience, the participants should be lit in a manner to allow the telepresence system to capture an accurate flesh tone.

Whilst floors are subject to the same considerations as wall with respect to reflected light, colour and sound transmission, there are also other physical considerations. Floors interact directly with participants and as a result create sound, for example, a chair rolling across a surface. Therefore, the floor material should also be chosen to minimise sound created through movement. Floors also need to provide a stable safe surface for any people, furniture or equipment in the room.

Ceilings are again subject to the same considerations as walls however given that most lighting is overhead the ceiling must be able to safely incorporate lighting.

10.2.3 Furniture

In a telepresence room, the tables and the chairs are arranged in a way that each participant in normal seating position falls in the capture range of only one camera and looks into that camera. To achieve a desirable experience, the curve of the tables should be aligned with the arrangement of cameras and screens.

It is suggested furniture captured by the cameras should have a shape or curve like the table to have consistency in the whole telepresence room. Also, furniture with heavy texture or complicated pattern may make the video coding more complex and the system less effective. It is suggested to place any additional furniture beyond capture areas and have clean and plain surfaced furniture with curves conforming to the capture areas. Decorations and adornments can be used to provide a sensation of depth, but their patterns and colours should be chosen carefully.

Furniture affects the acoustic and lighting properties of the room, in terms of sound reflection and light reflection, respectively.

In order to provide a more natural experience, peripheral equipment, such as computer cases, wireless routers and server cabinets are suggested to be concealed in large-sized furniture (desk, cupboard, etc.) or be integrated into the framework for screens and cameras . It is suggested that cables be hidden from the cameras, and best be embedded underneath the floor or floor covering. Table-mounted microphones are suggested to be flat (not with gooseneck), and it is suggested to hang ceiling-mounted microphones above the upper-boundary of captures.

To provide a better experience, these factors should be considered when designing a telepresence room or introducing changes to a telepresence room.

10.2.4 Lighting

For a pleasing telepresence experience, good lighting in terms of quantity and quality is needed for the environment where the telepresence endpoint is placed. A number of factors should be considered when designing and commissioning a telepresence room:

– Luminance distribution
– Illuminance
– Glare
– Directionality of light
– Colour aspect of the light and surfaces
– Flicker
– Daylight
– Maintenance

[ISO 8995-1] on the lighting of indoor work spaces provides further guidance on the above factors.

10.2.5 Acoustics

As discussed above, the size and shape of a room including the materials that the walls, floors and ceiling are made from affect the acoustic properties of a telepresence room. Important parameters to consider are reverberation time, ambient noise and sound insulation. Reverberation time should be kept to a minimum as sound quality decreases as reverberation time increases. Ambient noise should be minimized in order to provide a pleasing audio experience. Sound insulation is also important to minimize both the sound admitted to and transmitted from the telepresence room. Minimizing the admitted sound will increase the telepresence experience. Minimizing the transmitted sound will increase the privacy of the meeting.

10.2.6 Heating ventilation and air conditioning

As the main goal of a telepresence system is to provide a pleasing experience for the participants, a telepresence room should provide a comfortable environment for them. Heating ventilation and air conditioning (HVAC) systems include temperate control and ventilation. When designing these systems, there needs to be several considerations:

– People loading: The amount of people expected in the room needs to be considered. Each person generates heat and needs a certain flow of air to be comfortable.
– Equipment loading: Electronic equipment generates heat and needs to be cooled for efficient operation.
– Noise: The movement of air around obstacles creates noise which contributes to the ambient noise of the room. HVAC and equipment cooling systems should be designed to minimize noise.
– Airflow: Airflow is important to evenly distribute the temperature throughout the room. HVAC systems should be designed to evenly distribute the temperature without causing localized areas of temperature or airflow that may cause discomfort to participants. There also should be adequate airflow surrounding equipment to enable cooling.
– Thermal insulation: Due to acoustic treatments the walls, ceiling and floors in telepresence rooms may have a higher level of thermal insulation and less air leakage than in other parts of the building. This higher level of insulation and lower level of air leakage should be taken into account when designing the HVAC systems.
– Utilization time: Telepresence rooms may be occupied at different times of the day or night. Consideration should be given to a HVAC system that provides an optimum environment whilst the room is occupied and offers an energy saving mode when not.

10.2.7 Power requirements

Telepresence system hardware and telepresence room electrical equipment (i.e., lighting) should be designed to minimize energy usage where possible. This minimizes heat generation and resulting HVAC load which results in lower costs for energy consumption. Standby modes with low energy consumption should be considered for telepresence systems and rooms when the telepresence endpoint is not in use.

The power should be of sufficient quality to allow the operation of sensitive telepresence system and computer hardware.

When calculating the power required for a telepresence room, the following should be considered:

– The power consumption of the telepresence system hardware, including monitors, servers etc.
– The power consumption of supplementary devices such as projectors, extra displays.
– Power outlets associated with participants to allow them access to power for laptops, charging, etc.
– Lighting power consumption.
– Any local HVAC system requirements.
– Power related to supplementary uses, e.g., kettles, lamps.

Power outlets should be provided in close proximity to both the telepresence system hardware and, if required, close to the participants.

10.2.8 Network connectivity

Telepresence rooms are required to provide some form of network connectivity appropriate for the telepresence system deployed to it. The size, the audio and video quality of the telepresence and resulting bandwidth to transport media flows will largely dictate the type of network connection.

Consideration should also be given to providing network connectivity for the room participants so that they can connect laptops or other Internet-enabled devices.

Network connectivity may also be needed for HVAC devices, e.g., thermostats.

Connections to the PSTN may also be required.

Bibliography

[b-IETF RFC 5117] IETF RFC 5117 (2008), *RTP Topologies.*

[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