ETSI TR 126 923 V18.0.0 (2024-05)

Digital cellular telecommunications system (Phase 2+) (GSM);
Universal Mobile Telecommunications System (UMTS);
LTE;
5G;
Study on media handling aspects of IP Multimedia Subsystem (IMS) based telepresence
(3GPP TR 26.923 version 18.0.0 Release 18)

Reference
RTR/TSGS-0426923vi00

Keywords
5G,GSM,LTE,UMTS

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - APE 7112B
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° w061004871

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to normative deliverables may have been declared to ETSI. The declarations pertaining to these essential IPRs, if any, are publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (https://ipr.etsi.org/).

Pursuant to the ETSI Directives including the ETSI IPR Policy, no investigation regarding the essentiality of IPRs, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Trademarks

The present document may include trademarks and/or tradenames which are asserted and/or registered by their owners. ETSI claims no ownership of these except for any which are indicated as being the property of ETSI, and conveys no right to use or reproduce any trademark and/or tradename. Mention of those trademarks in the present document does not constitute an endorsement by ETSI of products, services or organizations associated with those trademarks.

DECT™, **PLUGTESTS™**, **UMTS™** and the ETSI logo are trademarks of ETSI registered for the benefit of its Members. **3GPP™** and **LTE™** are trademarks of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners. **oneM2M™** logo is a trademark of ETSI registered for the benefit of its Members and of the oneM2M Partners. **GSM**® and the GSM logo are trademarks registered and owned by the GSM Association.

Legal Notice

This Technical Report (TR) has been produced by ETSI 3rd Generation Partnership Project (3GPP).

The present document may refer to technical specifications or reports using their 3GPP identities. These shall be interpreted as being references to the corresponding ETSI deliverables.

The cross reference between 3GPP and ETSI identities can be found under https://webapp.etsi.org/key/queryform.asp.

Modal verbs terminology

In the present document "**should**", "**should not**", "**may**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the ETSI Drafting Rules (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

Contents

Foreword

The present document has been produced by the 3rd Generation Partnership Project (3GPP).

The contents of the present document are subject to continuing work within the TSG and may change following formal TSG approval. Should the TSG modify the contents of the present document, it will be re-released by the TSG with an identifying change of release date and an increase in version number as follows:

Version x.y.z

where:

- x the first digit:
  - 1 presented to TSG for information;
  - 2 presented to TSG for approval;
  - 3 or greater indicates TSG approved document under change control.
- y the second digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc.
- z the third digit is incremented when editorial only changes have been incorporated in the document.

Introduction

During Release 12, use cases and requirements on IMS-based telepresence were introduced by SA1 into TS 22.228 to enable telepresence support in IMS applications. In TS 22.228, telepresence is defined as a conference with interactive audio-visual communications experience between remote locations, where the users enjoy a strong sense of realism and presence between all participants (i.e. as if they are in same location) by optimizing a variety of attributes such as audio and video quality, eye contact, body language, spatial audio, coordinated environments and natural image size. A telepresence system is defined as 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) and environmental characteristics are used to establish telepresence.

The core network aspects of IMS-based telepresence have been addressed by the CT groups (CT1, CT3, CT4), including incorporation of new tools into IMS as defined by IETF's ControLling mUltiple streams for tElepresence (CLUE) WG (see more from: https://datatracker.ietf.org/wg/clue/charter/) that achieves media advertisement and configuration to facilitate controlling and negotiating multiple spatially related media streams in an IMS conference supporting telepresence, taking into account capability information, e.g. screen size, number of screens and cameras, codecs, etc., so that sending system, receiving system, or intermediate system can make decisions about transmitting, selecting, and rendering media streams.

This present document provides a study of media handling aspects of IMS-based telepresence in 3GPP services. This includes identification of codec-level technical gaps for a telepresence UE (TP UE), which is expected to not only support Multimedia Telephony Service for IMS (MTSI) UE media handling capabilities [2], but also more advanced media handling capabilities. Other SA4-level media handling aspects such as media configuration and session control, data transport, media adaptation, QoS handling and interworking with MTSI are also studied.

1 Scope

The present document provides a study on the media handling aspects of IMS-based telepresence in 3GPP services. This includes the investigation of the following areas:

- Media codecs (speech, video, real-time text) for IMS-based telepresence.
- Media configuration including session setup and control procedures for IMS-based telepresence, and media provisioning aspects of capability negotiation based on SDP and CLUE protocols, etc.
- Set-up and control of the individual media streams between clients including interactivity, such as adding and dropping of media components, as well as end-to-end QoS handling, etc. for IMS-based telepresence.
- Data transport including usage of RTP / RTCP protocols, RTP profiles, RTP payload formats, RTP mapping, media synchronization, etc. for IMS-based telepresence, e.g. in relation to negotiation and establishment of the CLUE data channel, and exchange of CLUE ADVERTISEMENT and CONFIGURE messages.
- Requirements and guidelines for media adaptation in IMS-based telepresence, for example in response to changes of network bandwidth.
- Media handling requirements and guidelines for fixed-mobile interworking as well as interworking with MTSI and with GSMA's IMS profile on High-Definition Video Conference (HDVC) service in IR.39.

The gap analysis of the above areas and associated recommendations and conclusions for the proposed improvements are documented in the present document. Study and evaluation of end-to-end quality of experience (QoE) for IMS-based telepresence use cases for various codec, media handling and QoS configurations are also presented.

2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non-specific.
- For a specific reference, subsequent revisions do not apply.
- For a non-specific reference, the latest version applies. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document *in the same Release as the present document*.

[1] 3GPP TR 21.905: "Vocabulary for 3GPP Specifications".

[2] 3GPP TS 26.114: "IP Multimedia Subsystem (IMS); Multimedia telephony; Media handling and interaction".

[3] 3GPP TS 22.228: "Service requirements for the Internet Protocol (IP) multimedia core network subsystem (IMS); Stage 1".

[4] 3GPP TS 24.229: "IP multimedia scall control protocol based on Session Initiation Protocol (SIP) and Session Description Protocol (SDP); Stage 3".

[5] 3GPP TS 24.147: "Conferencing using the IP Multimedia (IM) Core Network (CN) subsystem; Stage 3".

[6] 3GPP TS 24.103: "Telepresence using the IP Multimedia (IM) Core Network (CN) Subsystem (IMS); Stage 3".

[7] 3GPP TS 24.803: "Study on Telepresence using the IP Multimedia (IM) Core Network (CN) Subsystem (IMS); Stage 3".

[8] draft-ietf-clue-framework-22 (April 2015): "Framework for Telepresence Multi-Streams".

[9] draft-ietf-clue-datachannel-09 (March 2015): "CLUE Protocol Data Channel".

[10] draft-ietf-clue-signaling-05 (March 2015): "CLUE Signaling".

[11] draft-ietf-clue-data-model-schema-10 (June 2015): "An XML Schema for the CLUE data model".

[12] void.

[13] draft-ietf-clue-rtp-mapping-04 (March 2015): "Mapping RTP streams to CLUE media captures".

[14] draft-ietf-mmusic-sctp-sdp-14 (March 2015): "Stream Control Transmission Protocol (SCTP) – Based Media Transport in the Session Description Protocol (SDP)".

[15] IETF RFC 4574: "The Session Description Protocol (SDP) Label Attribute".

[16] GSMA PRD IR.39: "IMS Profile for High Definition Video Conference (HDVC) Service".

[17] Recommendation ITU-T H.239 (10/2014): "Role management and additional media channels for H.300-series terminals ".

[18] IETF RFC 4796: "The Session Description Protocol (SDP) Content Attribute".

[19] Recommendation ITU-T G.719 (06/2008): "Low-complexity, full-band audio coding for high-quality, conversational applications".

[20] ISO/IEC 14496-3:2009: "Information technology - Coding of audio-visual objects. Part 3: Audio".

[21] IETF RFC 5404: "RTP Payload Format for G.719".

[22] IETF RFC 3016: "RTP Payload Format for MPEG-4 Audio/Visual Streams".

[23] Recommendation ITU-T H.264 (02/2014): "Advanced video coding for generic audio-visual services" | ISO/IEC 14496-10:2005: "Information technology - Coding of audio-visual objects - Part 10: Advanced Video Coding".

[24] Recommendation ITU-T F.734 (10/2014): "Definitions, requirements, and use cases for Telepresence Systems".

[25] Recommendation ITU G.1091 (10/2014): "Quality of Experience requirements for telepresence services".

[26] IETF RFC 3264 (2002): "An Offer/Answer Model with the Session Description Protocol (SDP)", J. Rosenberg and H. Schulzrinne.

[27] draft-ietf-tsvwg-sctp-dtls-encaps-09 (January 2015): "DTLS Encapsulation of SCTP Packets".

[28] draft-ietf-mmusic-sdp-bundle-negotiation-23 (July 2015): "Negotiating Media Multiplexing Using the Session Description Protocol (SDP)".

[29] Recommendation ITU-T H.TPS-AV (02/2015): "Audio/video parameters for telepresence systems ".

[30] IETF RFC 6184 (2011): "RTP Payload Format for H.264 Video", Y.-K. Wang, R. Even, T. Kristensen, R. Jesup.

[31] Recommendation ITU-T H.241 (02/2012): "Extended video procedures and control signals for ITU-T H.300-series terminals".

[32] IETF RFC 6236 (2011): "Negotiation of Generic Image Attributes in the Session Description Protocol (SDP)", I. Johansson and K. Jung.

[33] Recommendation ITU-T H.245 (05/2011): "Control protocol for multimedia communication".

[34] IETF RFC 4566 (2006): "SDP: Session Description Protocol", M. Handley, V. Jacobson and C. Perkins.

[35] IETF RFC 6464: "A Real-time Transport Protocol (RTP) Header Extension for Client-to-Mixer Audio Level Indication".

[36] Recommendation ITU-T H.264 (04/2013): "Advanced video coding for generic audiovisual services".

[37] Recommendation ITU-T H.265 (04/2013): "High efficiency video coding".

[38] draft-ietf-mmusic-sdp-simulcast-01 (July 2015): "Using Simulcast in SDP and RTP Session".

[39] draft-ietf-mmusic-data-channel-sdpneg-03 (July 2015): "SDP-based Data Channel Negotiation".

[40] N10032, MPEG audio subgroup, "Report on the Verification Test of MPEG-4 Enhanced Low Delay AAC", 2008, Hannover.

[41] Recommendation ITU-T H.TPS-SIG (10/2014): "Signalling for telepresence-enabled conferencing".

[42] Recommendation ITU-T H.420 (10/2014): "Telepresence System Architecture".

[43] Recommendation ITU-T H.323 (12/2009): "Packet-based multimedia communication systems".

[44] 3GPP TR 26.952: "Codec for Enhanced Voice Services (EVS); Performance characterization".

3 Definitions and abbreviations

3.1 Definitions

For the purposes of the present document, the terms and definitions given in 3GPP TR 21.905 [1] and the following apply. A term defined in the present document takes precedence over the definition of the same term, if any, in 3GPP TR 21.905 [1].

Actual Size: A rendered figure in a display is the same size as if the person is in the room.

Conference: An IP multimedia session with two or more participants. Each conference has a "conference focus". A conference can be uniquely identified by a user. Examples for a conference could be a Telepresence or a multimedia game, in which the conference focus is located in a game server.

Eye Contact: Mutual direct-gaze between two persons communicating.

Gaze Awareness: 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.

Gaze Direction: The vector positioned along the visual axis, pointing from the fovea of the looker through the centre of the pupil to the looked at spot.

IM session: An IP multimedia (IM) session is a set of multimedia senders and receivers and the data streams flowing from senders to receivers. IP multimedia sessions are supported by the IP multimedia CN Subsystem and are enabled by IP connectivity bearers (e.g. GPRS as a bearer). A user may invoke concurrent IP multimedia sessions.

Telepresence: A conference with interactive audio-visual communications experience between remote locations, where the users enjoy a strong sense of realism and presence between all participants by optimizing a variety of attributes such as audio and video quality, eye contact, body language, spatial audio, coordinated environments and natural image size.

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) and environmental characteristics are used to establish Telepresence.

3.2 Abbreviations

For the purposes of the present document, the abbreviations given in 3GPP TR 21.905 [1] and the following apply. An abbreviation defined in the present document takes precedence over the definition of the same abbreviation, if any, in 3GPP TR 21.905 [1].

AAC-ELD	Advanced Audio Coding – Enhanced Low Delay
AAC-LD	Advanced Audio Coding – Low Delay
AMR	Adaptive Multi-Rate
AMR-WB	Adaptive Multi-Rate Wideband
AS	Application Server
AV	Advanced Video Coding
BFCP	Binary Floor Control Protocol
CBP	Constrained Baseline Profile
CHP	Constrained High Profile
CLUE	ControLling mUltiple streams for tElepresence
DTLS	Datagram Transport Layer Security
EVS	Enhanced Voice Services
FB	Fullband
GSMA	Groupe Speciale Mobile Association
HD	High Definition
HDVC	High Definition Video Conference
HEVC	High Efficiency Video Coding
ICE	Interactivity Connectivity Establishment
IETF	Internet Engineering Task Force
MCC	Multiple Content Capture
MPEG	Moving Picture Experts Group
MRFC	Multimedia Resource Function Controller
MRFP	Multimedia Resource Function Processor
MTSI	Multimedia Telephony Service for IMS
NAT	Network Address Translation
NB	Narrowband
RTCP	RTP Control Protocol
RTP	Real-time Transport Protocol
SCTP	Stream Control Transmission Protocol
SDP	Session Description Protocol
SWB	Super-wideband
TP	Telepresence
TP UE	TelePresence User Equipment
UHD	Ultra High Definition

4 Overview of IMS-based Telepresence in 3GPP

4.1 Introduction

During Release 12, IMS-based telepresence was introduced by SA1 in 3GPP services, and core network aspects were specified by the CT groups. This clause provides an overview of IMS-based telepresence in 3GPP.

4.2 Service Aspects

4.2.1 Overview

The use cases and requirements on IMS-based telepresence were introduced during Release 12 into TS 22.228 [3] to enable telepresence support in IMS applications.

In TS 22.228, telepresence is defined as a conference with interactive audio-visual communications experience between remote locations, where the users enjoy a strong sense of realism and presence between all participants (i.e. as if they are in same location) by optimizing a variety of attributes such as audio and video quality, eye contact, body language,

spatial audio, coordinated environments and natural image size. A telepresence system is defined as 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) and environmental characteristics are used to establish telepresence.

In order to provide a "being there" experience for conversational audio and video session between remote locations, where the users enjoy a strong sense of realism and presence, capabilities and preferences need to be co-ordinated and negotiated between local and remote participants such as:

- audio and video spatial composition information; e.g. spatial relationship of two or more objects (audio/video sources) in the same room to allow for accurate reproduction on the receiver side;
- capabilities of cameras, screens, microphones and loudspeakers, and their relative spatial relationships;
- meeting description, such as view information, language information, participant information, participant type, etc.

4.2.2 Use Case

In the scenario below depicted in Figure 4.1, a project team has one of their weekly reporting meetings using a Telepresence conference. Steven, John and Marc are in a meeting room in San Francisco. Fred and Liu are in another room in Paris. Ted is on the road and joins using his mobile phone. Bill is at home and joins using his PC. Both meeting rooms are equipped with multiple cameras and large screen monitors. Three cameras and screens are arranged to provide a panoramic view of the room. Additional cameras and screens are used to share presentations among the participants.

Multiple video streams are shared along with additional information (e.g. spatial information, video resolution, and environmental), so that the user experience is as if they are in same location. Audio information (e.g. spatial information) is exchanged to facilitate the rendering of the audio in accordance with the rendering of the video. Users in the meeting enjoy a strong sense of realism and presence between all participants.

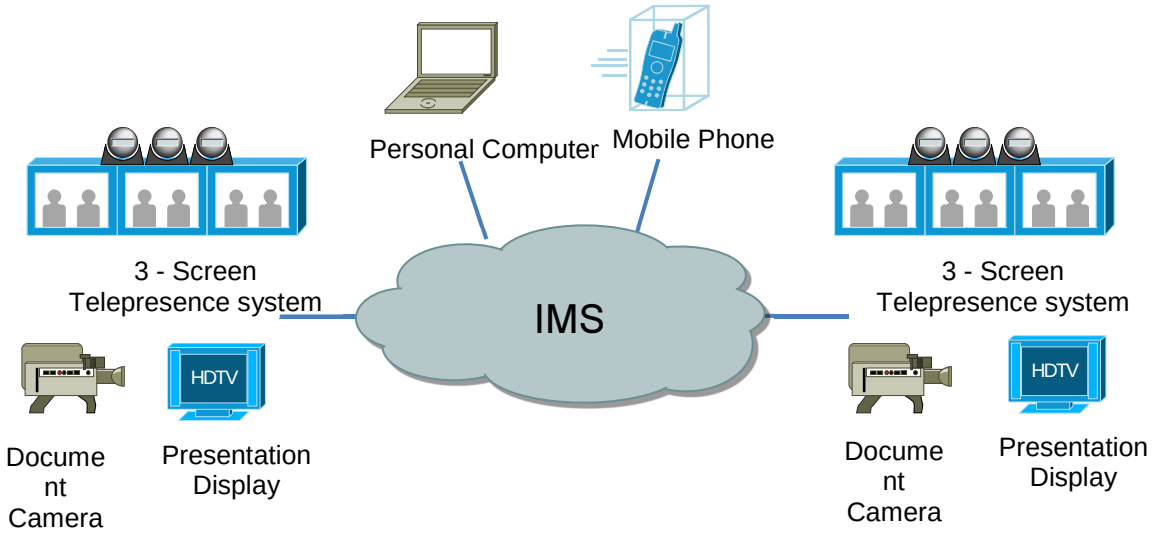

Figure 4.1: Telepresence

In the scenario depicted in Figure 4.2, participants may be from different operator's networks, or from enterprise networks. In such cases, IMS-based Telepresence has interconnection with Telepresence in other networks.

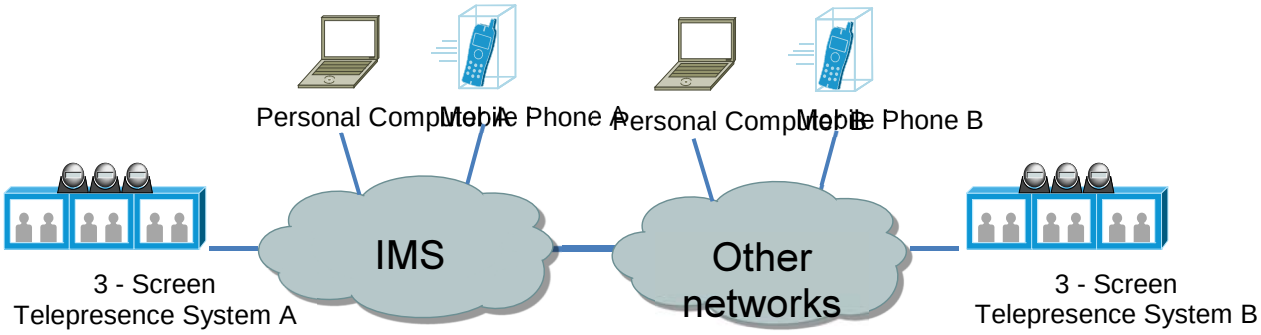

Figure 4.2: Interconnection of Telepresence

4.2.3 Requirements from TS 22.228

See clauses 7.10.2.2 and 8.8 of TS 22.228 [3].

4.3 Core Network Aspects

The core network aspects of IMS-based telepresence have been addressed by the CT groups, including incorporation of new tools into IMS as defined by IETF's ControLling mUltiple streams for tElepresence (CLUE) WG (see more from: https://datatracker.ietf.org/wg/clue/charter/) [8]-[13] that achieves media advertisement and configuration to facilitate controlling and negotiating multiple spatially related media streams in an IMS conference supporting telepresence, taking into account capability information, e.g. screen size, number of screens and cameras, codecs, etc., so that sending system, receiving system, or intermediate system can make decisions about transmitting, selecting, and rendering media streams. With the establishment of the CLUE data channel, the participants have consented to use the CLUE protocol mechanisms to determine the capabilities of the each of the endpoints with respect to multiple streams support, via the exchange of an XML-based data format. The exchange of CLUE messages of each participant's "advertisement" and "configure" is to achieve a common view of media components sent and received in the IM session supporting telepresence.

Enabling telepresence support involves updating and enhancing the existing IMS procedures for point-to-point calls as specified in 3GPP TS 24.229 [4] and for multiparty conferences as specified in 3GPP TS 24.147 [5]. This has been addressed in a new specification, TS 24.103 [6], which incorporates the CLUE framework with the Session Initiation Protocol (SIP), Session Description Protocol (SDP) and Binary Floor Control Protocol (BFCP) to facilitate controlling multiple spatially related media streams in an IM session supporting telepresence.

In [6], CT1 specifies procedures to deal with multiple spatially related media streams according to the CLUE framework to support telepresence and to interwork with IM session as below:

1) Initiation of telepresence using IMS, which includes an initial offer/answer exchange establishes a basic media session and a CLUE channel, CLUE exchanges to "advertisement" and "configure" media components used in the session, then followed by an SDP offer/answer in Re-INVITE request to complete the session establishment (see more for the general idea in draft-ietf-clue-framework [8]);

2) Release or leaving of an IM session supporting telepresence, which needs remove the corresponding CLUE channel;

3) Update of an ongoing IM session supporting telepresence, triggered by CLUE exchanges modifying existing CLUE information. For example: a new participant at an endpoint may require the establishment of a specific media stream;

4) Presentation during an IM session supporting telepresence, which may also be initiated by the exchange of CLUE messages and possibly need an updated SDP offer/answer and activation of BFCP for floor control; and

5) Interworking with normal IM session, this is to let the normal IMS users be able to join telepresence using IMS.

4.4 Protocols and Architecture for IMS-based Telepresence

4.4.1 Introduction

SIP, as specified in 3GPP TS 24.229 [4] and 3GPP TS 24.147 [5], is used as the basic session control protocol to create an IM session supporting telepresence.

The usage of SIP for the point to point call supporting telepresence follows the procedures as specified in 3GPP TS 24.229 [4]. The usage of SIP for the multiparty conference supporting telepresence follows the procedures as specified in 3GPP TS 24.147 [5].

To support an IM session supporting telepresence, a "+sip.clue" Contact header field parameter is defined in clause 3 of draft-ietf-clue-signaling [10].

SDP, as specified in 3GPP TS 24.229 [4] and 3GPP TS 24.147 [5], is used to establish multimedia streams in an IM session supporting telepresence. Each party in the session usually sends and receives multiple multimedia streams, which may not be symmetric due to their different capabilities for media production and rendering.

CLUE, as specified in draft-ietf-clue-framework [8], is used to advertise and configure audio and video components comprising the media flows in an IM session supporting telepresence. A data channel for CLUE message, as defined in draft-ietf-clue-datachannel [9], is negotiated via the first INVITE message when creating an IM session supporting telepresence. With the establishment of that channel, the participants have consented to use the CLUE protocol mechanisms to determine the capabilities of the each of the endpoints with respect to multiple streams support. The following exchange of CLUE messages of each participant's "advertisement" and "configure" is to achieve a common view of media components sent and received in the IM session supporting telepresence. A corresponding SDP offer/answer may be needed to establish the media streams based on the user's choice in CLUE messages.

BFCP, as specified in 3GPP TS 24.147 [5], is used to offer floor control of shared resources in an IM session supporting telepresence.

4.4.2 Functional Entities

4.4.2.1 General

As specified in TS 24.103 [6], the functional entities of an IMS-based telepresence system are depicted in Figure 4.3.

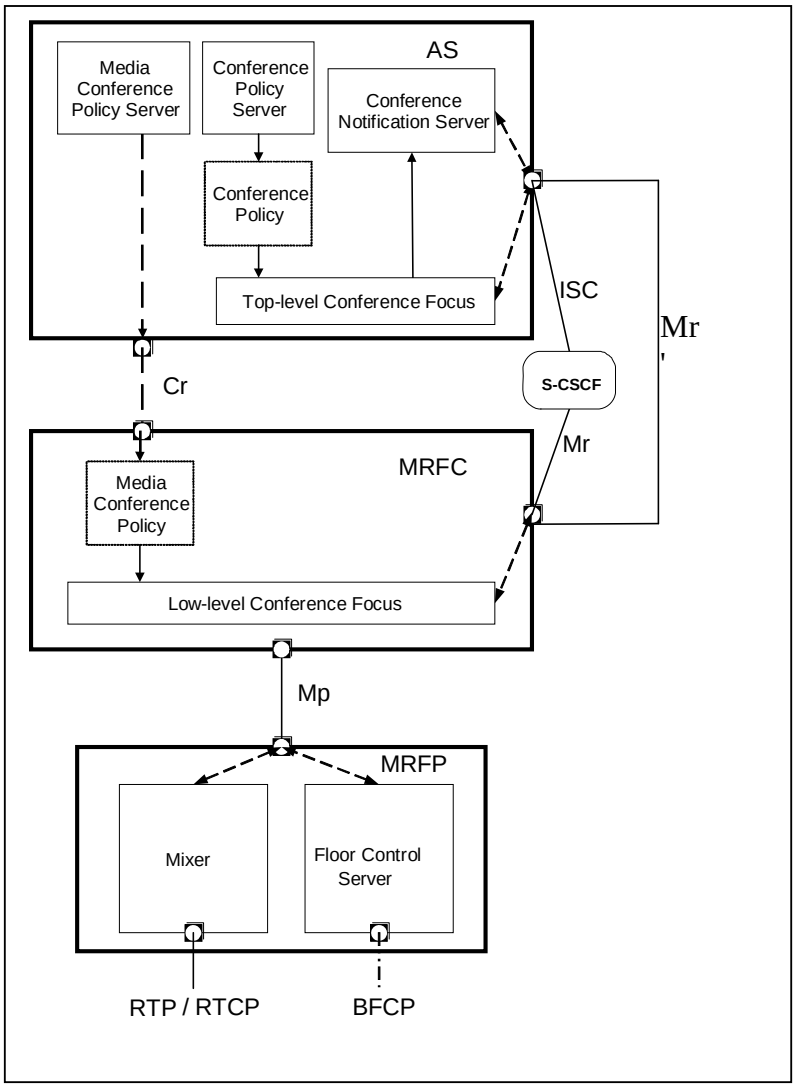

Figure 4.3: Functional Entities of an IMS-based Telepresence System

4.4.2.2 User Equipment (UE)

For the purpose of IM session supporting telepresence, the TP UE implements the role of a participant in both point to point call and multiparty conference. As such, during registration the TP UE follows the procedures specified in subclause 5.1.1 of 3GPP TS 24.229 [4] as a basis, and it includes a "+sip.clue" Contact header field parameter in the SIP REGISTER request.

To establish an IM session supporting telepresence, the TP UE uses the procedures as specified in subclause 5.1 of 3GPP TS 24.229 [4] for point to point call, and follows the conference procedures as specified in subclause 5.3.1 of 3GPP TS 24.147 [5] for multiparty conference, the TP UE includes a "+sip.clue" Contact header field parameter in the SIP INVITE and SIP UPDATE requests and responses.

TP UE supports the point to point call procedures specified in subclause 6.1 of 3GPP TS 24.229 [4], and also the conference procedures specified in subclause 6.3.1 of 3GPP TS 24.147 [5] appropriate to the functional entity in which the participant is implemented for multiparty conference.

As described in TS 24.103 [6], to initiate an IM session supporting telepresence, the TP UE executes the following procedures:

1) generate an SDP offer in the SIP request, containing basic media streams and an establishment request for a DTLS/SCTP association used to realize a CLUE data channel. The initial SDP offer message negotiates the port and transport information for setting up the DTLS/SCTP association, via a separate SDP "m=" line together with SDP "fmtp" and "sctp-port" attribute that indicating the use of a data channel, and a further session-level "group" attribute may be carried to indicate its belonging of a CLUE group, which also contains all the CLUE-controlled media. The procedures for establishment of the DTLS/SCTP association via SDP can be found in draft-ietf-mmusic-sctp-sdp [14].

2) set up the DTLS/SCTP association used for a CLUE data channel with the remote party, after receiving the SDP answer with information for CLUE data channel establishment (e.g. an SDP "m=" line together with the SDP "fmtp" and "sctp-port" attributes to describe a DTLS/SCTP association indicating the use of a data channel).

3) When the DTLS/SCTP association used to realize a CLUE data channel is established, the TP UE opens the CLUE data channel via a DCEP message based on a SCTP stream. After receiving the response DCEP message, the CLUE data channel is successfully opened. The procedures for using such a protocol can be found in clause 4 of draft-ietf-clue-datachannel [9]. As defined in draft-ietf-clue-signaling [10] and draft-ietf-clue-datachannel [9], only a single CLUE data channel is established within the same IM session supporting telepresence.

Once the CLUE data channel is connected, the TP UE triggers an updated SDP offer/answer exchange to set up media streams for transmission of the media captures. When constructing the SDP offer in both of point to point call and multiparty conference, the TP UE executes the following procedures:

1) add a corresponding "m=" line for each encoding defined in CLUE messages, using an associated "label" attribute for each "m=" line to identify the each encoding in CLUE messages. In addition, mark the "m=" line(s) as send only with an "a=sendonly" attribute or as inactive with an "a=inactive" attribute used to represent the sender's encode ability and policies; and

2) use an SDP "group" session-level attribute to signal that the "m=" line(s) are CLUE-controlled.

When the SDP offer is sent, the following SDP negotiation procedures follows normal rules as defined in IETF RFC 3264 [26].

In the conference case, if a TP UE acting as a conference participant joins in the conference, the TP UE follows the above steps to set up an IM session supporting telepresence.

Further detail on session update and termination procedures for a TP UE can be found in TS 24.103 [6].

The TP UE typically supports both of the two roles in CLUE protocol, i.e. media provider and media consumer, in an IM session supporting telepresence. However, a TP UE may only act as a media provider if it only sends media streams but does not receive, or it may only act as a media consumer if it only receives media streams.

As specified in draft-ietf-clue-framework [8], a media provider representing an entity that intends to send CLUE-controlled media streams, sends CLUE ADVERTISEMENT message to describe the content of the media and the media streams encodings via encoding groups and individual encodings as specified in clause 9 of draft-ietf-clue-framework [8]. As such, a TP UE acting as a media provider sends a CLUE ADVERTISEMENT message and receives the corresponding CLUE CONFIGURE message from the remote party. The detailed information and format of CLUE messages can be found in draft-ietf-clue-data-model-schema [11].

Once the CLUE data channel is connected during an IM session supporting telepresence, the TP UE, acting as a media provider sends a CLUE ADVERTISEMENT message with multiple media captures related to the video and audio content it can provide, and carry the identities of these media captures together with information elements to describe characteristics of these media captures including video capture spatial information in order to allow the remote party to select the source(s) it wants to consume, according to draft-ietf-clue-framework [8].

The TP UE, acting as media provider, may send a new CLUE ADVERTISEMENT message to update the media captures anytime during the ongoing session, as specified in draft-ietf-clue-framework [8]. When the media provider receives the CLUE CONFIGURE message, the media provider triggers an updated SDP offer to establish the media streams based on the latest received CLUE CONFIGURE message.

To provide media switching or composition for multiple contents with respect to time and space during an IM session supporting telepresence, the acting media provider, a TP UE can include one or more Multiple Content Capture (MCC) that is composed of multiple individual captures in the CLUE ADVERTISEMENT message, as defined in subclause 7.2 of draft-ietf-clue-framework [8].

As specified in draft-ietf-clue-framework [8], a media consumer representing an entity that intends to receive CLUE-controlled media streams, sends CLUE CONFIGURE message to specify the content and media streams it wants to receive. A TP UE acting as a media consumer can send CLUE CONFIGURE message to select the media captures based on the latest received CLUE ADVERTISEMENT message. The detailed information and format of CLUE messages can be found in draft-ietf-clue-data-model-schema [11].

Once the CLUE data channel is connected in both point to point cases and conference cases, the TP UE acting as a media consumer executes the following:

1) be prepared to receive a CLUE ADVERTISEMENT message from a remote party;

2) select the media captures it wants to receive based on the information elements of the media captures in the received CLUE ADVERTISEMENT message; and

3) send the selected media captures within a CLUE CONFIGURE message to the remote party.

The TP UE, acting as a media consumer, may choose to have a switching or composed media by using of Multiple Content Capture (MCC) via CLUE exchange as defined in subclause 7.2 of draft-ietf-clue-framework [8].

4.4.2.3 Media Resource Function Controller (MRFC) and Media Resource Function Processor (MFRP)

For the purpose of IM session supporting telepresence, the MRFC supports the procedures as described in subclauses 5.2.2 and 6.2.2 of 3GPP TS 24.147 [5].

The MRFC and MRFP support both of the two roles in CLUE protocol, i.e. media provider and media consumer, as specified in draft-ietf-clue-framework [4].

4.4.2.4 Conferencing Application Server (Conferencing AS)

For the purpose of IM session supporting telepresence, the TP enabled conferencing AS implements the role of a conference focus. The TP conferencing AS may also implement the role of a participant.

The TP conference focus follows the procedures specified in subclauses 5.3.2 and 6.3.2 of 3GPP TS 24.147 [5] appropriate to the functional entity in which the conference focus is implemented. The TP conference focus includes the "+sip.clue" Contact header field parameter when generating a SIP request or response for the establishment of an IM conference supporting telepresence.

Once CLUE data channels between the TP enabled conference focus and TP UEs are connected, the TP enabled conference focus executes the following:

1) trigger an updated SDP offer/answer exchange to set up media streams for transmission of the media captures between individual TP UE and the TP enabled conference focus based on the received CLUE CONFIGURE messages; and

2) follow the procedures with TP UEs as described in subclause 4.2.2.2.

When the conference focus receives an update message to the ongoing IM session supporting telepresence, the conference focus executes the following:

1) finalize the CLUE exchange and possible update of SDP offer/answer according to clause 5 in draft-ietf-clue-signaling [10]; and

2) if session update is needed between the TP enabled conference focus and other participants in the same conference, the TP enabled conference focus initiates the update procedures accordingly.

The Conferencing AS supports both of the two roles in CLUE protocol, i.e. media provider and media consumer, as specified in draft-ietf-clue-framework [8].

A TP-enabled conference focus, acting as a media provider, sends a CLUE ADVERTISEMENT message and receives the corresponding CLUE CONFIGURE message from the remote party. The detailed information and format of CLUE messages can be found in draft-ietf-clue-data-model-schema [11].

Once the CLUE data channel is connected during an IM session supporting telepresence, the TP enabled conference focus, acting as a media provider sends a CLUE ADVERTISEMENT message with multiple media captures related to the video and audio content it can provide, and carry the identities of these media captures together with information elements to describe characteristics (see clause 4) of these media captures including video capture spatial information in order to allow the remote party to select the source(s) it wants to consume, according to draft-ietf-clue-framework [8].

The TP enabled conference focus acting as a media provider may further construct a new CLUE ADVERTISEMENT message to a TP UE in the conference based on the media capture information received from the other TP UEs.

To provide media switching or composition for multiple contents with respect to time and space during an IM session supporting telepresence, the acting media provider, a TP enabled conference focus, can include one or more Multiple Content Capture (MCC) that is composed of multiple individual captures in the CLUE ADVERTISEMENT message, as defined in subclause 7.2 of draft-ietf-clue-framework [8].

A TP enabled conference focus acting as a media consumer sends CLUE CONFIGURE message to select the media captures based on the latest received CLUE ADVERTISEMENT message. The detailed information and format of CLUE messages can be found in draft-ietf-clue-data-model-schema [11].

The TP enabled conference focus, acting as a media consumer, may choose to have a switching or composed media by using of Multiple Content Capture (MCC) via CLUE exchange as defined in subclause 7.2 of draft-ietf-clue-framework [8].

In the conference case, the TP enabled conference focus can select the media captures based on the CLUE CONFIGURE messages received from other TP UEs.

5 Gap Analysis on Media Handling Aspects of IMS-based Telepresence

5.1 Media Handling Aspects of CLUE

5.1.1 Introduction on CLUE

CLUE, under development in IETF, is used to advertise and configure audio and video components comprising the media flows in an IM session supporting telepresence, see more from: https://datatracker.ietf.org/wg/clue/charter/ [8] - [13]. A data channel for CLUE message is negotiated via the first INVITE message when creating an IM session supporting telepresence. With the establishment of that channel, the participants have consented to use the CLUE protocol mechanisms to determine the capabilities of the each of the endpoints with respect to multi-stream support, via the exchange of an XML-based data format. The following exchange of CLUE messages of each participant's "advertisement" and "configure" is to achieve a common view of media components sent and received in the IM session supporting telepresence. A corresponding SDP offer/answer may be needed to establish the media streams based on the user's choice in CLUE messages.

A Telepresence (TP) UE is expected to support CLUE while an MTSI UE is not, and therefore all media handling aspects relevant for enabling or enhancing CLUE support are relevant for this study.

While TS 24.103 [6] addresses IMS aspects of telepresence at a core network level, 3GPP-based media handling requirements and features for a TP UE with regards to the usage of CLUE have not been established. The following gaps are observed.

5.1.2 Media Handling Requirements for TP UE

5.1.2.1 Data Channel for CLUE Messages

The exchange of CLUE messages requires a data transport channel over DTLS/SCTP (Datagram Transport Layer Security / Stream Control Transmission Protocol) [14] negotiated via the initial SDP offer and answer, and usage of the SDP-based "SCTP over DTLS" data channel negotiation mechanism (see more in draft-ietf-mmusic-data-channel-sdpneg [39]) in order to open the CLUE data channel based on a SCTP stream in each direction. Therefore a TP UE needs to support these protocols over the user plane, while the MTSI client protocol stack depicted Figure 4.3 of TS 26.114 currently lacks these capabilities.

The non-media data conveyed over a data channel is handled by using SCTP encapsulated in DTLS. Furthermore, Using DTLS over UDP in combination with ICE enables middle box traversal in IPv4 and IPv6 based networks. This data transport service operates in parallel to the RTP media transport, and all of them can be eventually share a single transport-layer port number.

The layering of protocols for data channel is shown in the following Figure 5.1.

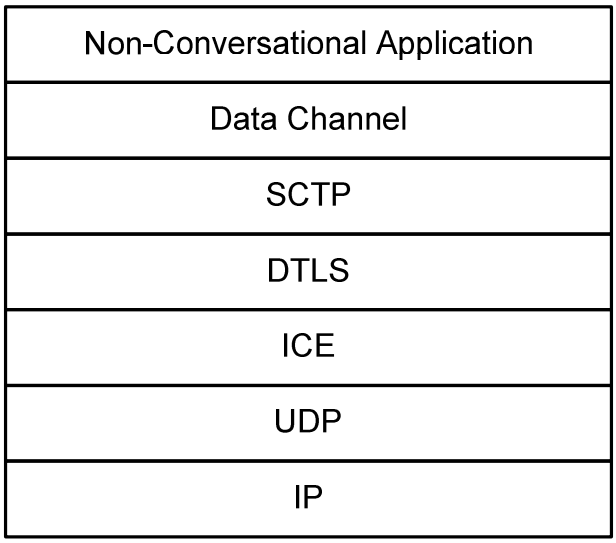

Figure 5.1: Protocol stack of data channel

The session setup for data channel transported non-media data can determine: IP address, UDP port number, SCTP port number (the default value is 5000), SCTP protocol identifier (i.e. DTLS/SCTP), media format and additional session parameters.

A TP UE can offer a DTLS/SCTP association together with the media format indicating the use of a data channel in the first SDP offer or subsequent SDP offers. A TP client can further open the data channel via the SDP-based "SCTP over DTLS" data channel negotiation mechanism to indicate specific non-conversational application (e.g. CLUE protocol) over it.

5.1.2.2 RTP / RTCP Level Requirements

No media handling requirements and recommendations on a TP UE have been specified in TS 24.103 with regards to the usage of RTP / RTCP protocols in relation to negotiation and establishment of the CLUE data channel. It is expected that the normative work will address these gaps.

One such example media handling aspect is the use of RTP multiplexing and mapping of RTP streams to CLUE media captures in the context of IMS-based telepresence. Due to the potentially large number of RTP flows required for a TP session involving potentially many endpoints, each of which can have many media captures and media renderers, it is desirable to multiplex multiple RTP media flows onto the same transport address, so to avoid using the port number as a multiplexing point and the associated shortcomings such as NAT/firewall traversal. The mapping of those RTP flows to the header fields of the RTP packets incurs the large number of possible permutations, and hence it can be beneficial to enable a mapping framework that allows narrowing down the number of possible options that a SIP offer-answer exchange has to consider. Such mappings include CLUE individual encodings to SDP and RTP media streams to SDP.

Every m-line representing CLUE encoding contains a "label" attribute as defined in RFC 4574 [15]. Various recommendations are provided in [13] addressing how RTP and RTCP streams should be encoded and transmitted, and how their relation to CLUE Media Captures should be communicated, and can also be considered for a TP UE in the context of IMS-based telepresence.

Another example problem to be observed in this context has already been observed in clause 5.8 of TR 24.803 [7], with regards to the creation of RTP streams before the completion of CLUE negotiations. Various solution approaches have also been analysed, such as the following:

- No media exchanged until completion of CLUE negotiation, i.e. hold the basic RTP streams until the completion of CLUE negotiation. Once the CLUE negotiation is completed, SDP may or may not be updated to adapt the transport of CLUE encodings.

- Transfer simple media content, where the initial SDP offer-answer is negotiated with the basic RTP streams, which are configured based on a simple one-screen/one-speaker view of the room. The RTP streams are flowing on after they are established, even though the CLUE negotiation is still in progress. Once the CLUE negotiation is completed, SDP may or may not be updated to adapt the transport of CLUE encodings.

The information about CLUE-controlled media streams is split between SDP and CLUE, and CLUE-controlled media is not to be sent unless it has been negotiated both in CLUE and SDP based on the recommendations specified in clause 5 of the draft-ietf-clue-signaling [10]. In addition, an ongoing SDP exchange is not to be delayed regardless of the completion of CLUE negotiations.

In IMS-based telepresence, RTP streams and CLUE media captures can be initiated by the TP UEs and TP-enabled conference focus. Each RTP stream uses a single 5-tuple, and each CLUE media capture is a source of media (e.g. from one or more capture devices).

In order to associate the media in different protocols (i.e. RTP/RTCP, SDP and CLUE), the mapping of RTP streams to CLUE media captures can be achieved by the following steps:

1) a media capture in CLUE "configure" message is assigned respectively a specific media capture encoding (i.e. "encID"); and

2) the media capture encoding is represented by a CLUE-controlled m-line with a label value corresponding to the context of the "encID"; and

3) the RTP stream associated with the media capture can be identified by the transport address in the CLUE-controlled m-line associated with the media capture encoding.

If multiple RTP streams share a single 5-tuple using RTP multiplexing according to the draft-ietf-mmusic-sdp-bundle-negotiation [28], it is necessary to add an additional mechanism to identify these RTP streams except transport address. In order to realize the mapping, the CLUE-controlled m-line further contains an "mid" attribute, which is carried as both an RTP header extension and an RTCP SDES message.

Furthermore, if a media capture acts as an MCC (i.e. an Multiple Content Capture is composed of multiple other media captures) using media switching and composition as specified in subclauses 7.3.1.2.1 and 7.3.2.2.1 of 3GPP TS 24.103 [6], a media capture encoding can be assigned to the MCC in CLUE "configure" message. And therefore it needs to identify these media captures in the same MCC. Accordingly, the MCC specifies an identifier (i.e. CaptureId) for each media capture in it, and then the identifier(s) are carried as both an RTP header extension and an RTCP SDES message.

5.1.3 Guidelines and Examples of SDP and CLUE for Telepresence

In the MTSI context, TS 26.114 provides an extensive set of SDP examples as guidelines to facilitate vendors' implementations and ensure interoperability. A similar effort is desired for the anticipated normative work on IMS-based telepresence systems, where guidelines and examples of SDP and CLUE can be provided for various scenarios such as the following:

- Interactions of the CLUE protocol, with SDP offer/answer negotiations.

- SDP and CLUE examples that demonstrate various standardized features of the CLUE protocol as described in [8], such as grouping, SCTP/DTLS data channel activation, multiplexing of CLUE controlled media, etc.

- Sessions initiated by TP UE or initiated by media gateway.

- SDP and CLUE examples when adding, modifying or removing media components, as well as enabling and disabling CLUE during mid-call.
- SDP and CLUE examples for inter-working with other IMS or non-IMS IP networks, including MTSI.

5.1.4 Session Setup and Control Procedures for TP UE

The session set up for real-media data based on RTP is specified in subclause 6.2 of 3GPP TS 26.114 [2], subclauses 5.3.1 and 6.3.1 of 3GPP TS 24.103 [6]. Due to the asymmetric nature of CLUE, the negotiation of media configuration (i.e. bandwidth and QoS parameters) of the CLUE-controlled media should be unidirectional.

This session set-up for non-real-media (i.e. CLUE messages) based on DTLS/SCTP is specified in subclause 5.1.2.1. And the detail information about how SCTP is used on the top of the DTLS protocol can be found in draft-ietf-tsvwg-sctp-dtls-encaps [27].

A TP UE supporting multiple media components initiates a telepresence session, where CLUE-controlled media components may be complementary with each other (e.g. an additional audio stream is a translation or commentary of the conference). In order to associate the different CLUE-controlled media components, the TP UE should use the CLUE media capture attribute (i.e. "Related to") defined in draft-ietf-clue-framework [8].

During an IM session supporting telepresence, the session can be renegotiated due to adding and removing CLUE-controlled media based on subclause 6.3 of 3GPP TS 26.114 [2]. In particular, when adding CLUE-controlled media, the renegotiation of the session requires changes to both SDP and CLUE messages as specified in the clause 5 of draft-ietf-clue-signaling [10].

A TP UE supporting multiple CLUE-controlled media components may add one or more media components to an on-going telepresence session. In order to identify the CLUE-controlled media components with the same media type, the TP UE should use the CLUE media captures attributes (i.e. "presentation", "view", "description" and "lang") defined in draft-ietf-clue-framework [8].

5.1.5 Gap Analysis

The observed gaps from a media handling perspective include the following:

- User plane protocol stack for MTSI UEs does not include support of CLUE data channel
- Media handling aspects with regards to the usage of RTP / RTCP protocols need to be specified for TP UEs, e.g. the use of RTP multiplexing and mapping of RTP streams to CLUE media captures
- Guidelines and examples on SDP and CLUE usage should be provided for IMS-based telepresence.

5.2 GSMA IR.39 IMS Profile for High Definition Video Conference

5.2.1 Introduction

The IMS Profile for High Definition Video Conference (HDVC) service, documented in the specification IR.39 [16] defines a minimum mandatory set of features that a video communication client and the network are required to implement to guarantee an interoperable, high quality IMS-based video communication service over fixed and mobile access. The HDVC service comprise point-to-point video calls and video conferences with one full duplex audio stream with tight synchronization to one main video stream and another video stream aimed for sharing of for example presentation slides.

Several mandated or recommended media handling features of GSMA IR.39 could be adopted by a Telepresence (TP) UE toward better interoperating with an HDVC UE. In particular, the following UE capabilities mandated or recommended in the present document constitute a gap with respect to the media handling capabilities of an MTSI UE.

5.2.2 Screen Sharing

It is mandatory for an HDVC UE to support screen sharing by a dedicated video stream, typically using low frame rate and high resolution, similar to what is specified for ITU-T Recommendation H.239 [17].

The screen sharing media is sent from one sender to all participants in a conference.

The screen sharing media stream is identified in SDP with a video media line placed after the media line for the main video content. The media line uses the attribute "a=content:slides", as defined in IETF RFC 4796 [18] and uses the "3gpp_sync_info: No Sync" attribute as defined in clause 6.2.6 of 3GPP TS 26.114 [2] to indicate that the stream is not synchronized with the main voice and video streams.

Accordingly, it is mandatory for an HDVC UE to support the use of a second video stream in point to point video calls and in conferences.

5.2.3 Voice Codecs

Mandatory and recommended voice codec requirements of IR.39 for HDVC UEs contain those in TS 26.114 for MTSI clients using 3GPP access and partially those for MTSI clients using fixed access (there are some additional codecs recommended in TS 26.114 for MTSI clients using fixed access based on TS 181 005). Beyond that, some further requirements on voice codecs are specified in GSMA IR.39 for HDVC UEs. The discussion below provides some of these requirements.

If super-wideband or fullband voice is supported, an HDVC UE of GSMA IR.39 is required to support the Enhanced Voice Services (EVS) codec and recommended to support the ITU-T G.719 codec (described in ITU-T G.719 recommendation) [19] or the AAC-LD codec (described in recommendation ISO/IEC 14496-3:2009) [20]. Furthermore, according to the GSMA IR.39 profile, the entities in the IMS core network that terminate the user plane are mandated support the EVS codec. The ITU-T G.719 codec or the AAC-LD codec and the transcoding between EVS and these codecs may be supported if needed for super-wideband or fullband voice interoperability with UEs not supporting EVS (non HDVC UEs or HDVC UEs before EVS is introduced). Accordingly, the following RTP payload format considerations apply:

- G.719 Payload Format Considerations: When an HDVC UE supports G.719, it also supports the G.719 payload format of IETF RFC 5404 [21] according to GSMA IR.39.

- AAC-LD Payload Format Considerations: When an HDVC UE supports AAC-LD, it also supports the MP4A-LATM payload format of IETF RFC 3016 [22], according to GSMA IR.39.

Currently existing telepresence systems offer improved user experience over regular video conferencing. In order to provide a "being there" experience where the users enjoy a strong sense of realism and presence for conversational audio, super-wideband (SWB) and fullband (FB) audio coding can be considered for IMS-based telepresence services.

As an example, one of the codecs supporting SWB and FB audio coding is EVS. To show the quality benefits of SWB audio coding relative to WB and NB, 3GPP has conducted extensive EVS characterization testing, in TR 26.952 [44], where the following was concluded:

- Clean and Noisy speech: The EVS-SWB codec performance is significantly better than the previously standardized AMR-WB codec. The subjective quality of EVS-SWB coding at 9.6 kbps is better than the AMR-WB at 23.85 kbps. Further, the EVS-SWB codec performance is significantly better than the existing super-wideband codecs, such as the ITU-T G.719, both in clean channel as well as under impaired channel conditions. The EVS-SWB clean speech quality at 13.2 kbps is comparable to that of the ITU-T G.719 at 48 kbps.

- Music/audio coding: The subjective quality of EVS-SWB coding starting at 13.2 kbps is significantly better than the AMR-WB at its highest bit rate 23.85 kbps. Further, EVS-SWB shows major improvements for mixed-content and music, performing equally or better than the existing super-wideband codecs, such as the AMR-WB+, the ITU-T G.719 at much lower algorithmic delay than those codecs.

Consequently, super-wideband or fullband coding can be a more suitable and compelling choice for IMS-based telepresence services as the extended coding of audio bandwidth (as depicted in Figure 5.2), would provide not only an improved sense of presence and contribute to better music quality but also improve naturalness.

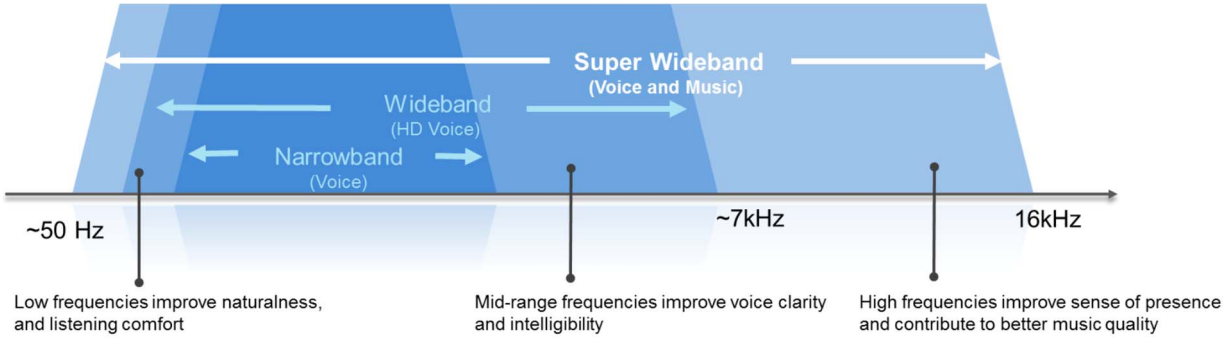

Figure 5.2: Audio bandwidth depiction, NB (~300 Hz to 4 kHz), WB (~100 Hz to 7 kHz), and SWB (~50Hz to 16 kHz)

5.2.4 Video Codecs

Mandatory and recommended video codec requirements of IR.39 for HDVC UEs contain those in TS 26.114 for MTSI clients, but also some further aspects on video codecs are specified in GSMA IR.39. In particular, GSMA IR.39 specifies optional support for H.264 Main and High Profiles [23] for the HDVC UE and the network.

5.2.5 Gap Analysis

The observed gaps from a media handling perspective include the following:

- Additional codec capabilities for speech and video that are lacking in MTSI UEs, e.g. those documented in GSMA IR.39 HDVC profile, may be considered for TP UEs.
- Further media handling aspects from GSMA IR.39 HDVC profile such as screen sharing may be considered for TP UEs.

5.3 Media Handling Aspects of Telepresence from ITU-T SG16

5.3.1 Functional Requirements

In [24], the functional requirements for telepresence systems are defined. Moreover, in [25] recommendations and guidelines for achieving high QoE in Telepresence services are provided. From a media handling perspective, the following requirements and recommendations are relevant and not covered by those in TS 22.228 [3]. Hence, 3GPP SA4 may consider these requirements and recommendations during the course of normative work on media handling aspects of IMS-based telepresence. The control-plane specific requirements from [24] addressed by TS 24.103 [6] based on IETF CLUE protocol are not addressed here.

5.3.1.1 User Experience

- It is recommended that a telepresence system support actual size high definition imaging of videoconferencing participants at the remote end. The support of actual size is considered an advanced telepresence feature.
- It is recommended that a telepresence system provide an experience where there is awareness of both direct gaze and averted gaze, with the remote participants' video images being rendered within a tolerable angle in both horizontal and vertical axes from the gaze direction between the two communicating participants. NOTE – The visual clues for determining the gaze direction may include the orientation of the looker's eyes with respect to their head and the orientation of the looker's head with respect to the environment. Some research suggests that tolerable angles are 3-5 degrees horizontal and up to 12 degrees upward and 8 degrees downward.
- It is recommended that a telepresence system support a means of locating the speaker in a remote meeting room so that local participants can readily identify the position of different remote speakers.

- It is recommended that for Telepresence systems: end to end video delay: ≤ 320 milliseconds, and end to end audio delay: ≤ 280 milliseconds. It is recommended that a telepresence system enable synchronization of audio and video from the remote location. Synchronization includes "lip-sync" and the sequencing of multiple media streams. It is recommended that audio streams should be less than 40 ms ahead of video streams, and less than 60 ms behind video streams.

- It is recommended that a telepresence system support multiple video sources in each meeting room. It is recommended that these video sources provide a seamless view of the conference participants.

- It is recommended that a telepresence system support multiple screens in each meeting room. It is recommended that the gap width between neighbouring displays be minimized.

- It is recommended that a telepresence system allow the receiver to switch captures between displays and assign desirable display positions for them (either automatically or via user input).

- It is recommended that a telepresence system indicate detailed information associated with a composed capture including the list of originating sources and the policy of the composition.

- [A telepresence system can optionally provide each participant with an experience of being present in the same virtual room, maintaining consistent spatial relationships among all the participants.]

5.3.1.2 Statistical Information Reporting

It is recommended that a telepresence management system be able to provide various statistical information about a telepresence conference session (both during and at the end of the conference) which includes but is not limited to: media type/s used, the numbers of endpoints, session duration, and data volume of each media type. Depending on the information required the telepresence management system may collect the information from multiple different telepresence system entities, e.g. call resource controller, multi-point control unit and endpoints.

It is recommended that a telepresence system provide QoS/QoE information including media network utilization, network delay, loss rate and jitter.

5.3.1.3 Network-Level Aspects

It is recommended that a telepresence system enable reservation of network resources for assuring QoS for the telepresence session.

It is recommended that a telepresence system enable adaptation of the quality of media based on the network status. For example: Under a best-effort network such as the Internet which cannot provide strictly guaranteed network QoS, the telepresence system may need to manipulate its video quality based on network status.

It is recommended that a telepresence system provide a means for endpoints to participate in telepresence conferences when some of telepresence system components are located in one or more private networks.

5.3.2 Audio/Video Parameters

5.3.2.0 General

In [41] and [42], ITU-T SG16 defines signalling and architecture for telepresence enabled conferencing. The ITU-T recommendation in [41] describes signalling procedures related to point-to-point and multipoint call processes in H.245 [33] based telepresence systems, including registration, call session establishment, capability exchange and negotiation of media stream transport. The ITU-T recommendation in [42] describes the telepresence system functions, devices and network elements for capturing, delivering, managing and rendering multiple high quality interactive audio and video signals in a telepresence conference. The CLUE framework is utilized as part of ITU-T's signalling and architecture for telepresence enabled conferencing in order to signal information related to multiple media captures. As such, a H.323 telepresence session follows additional call signalling procedures to negotiate the use of CLUE signalling and to link CLUE information to H.245 capabilities.

In this context, ITU-T SG16 also defines in [29] the audio/video/environment endpoint parameters for telepresence systems applicable for both SIP/IMS [4]-[5] and H.323 [43] based systems, including those which are used in the media setup phase during capability negotiation, and [29] also gives the specific format definitions for these parameters.

Collectively, these audio/video parameters and their associated values can be expected to provide a high quality telepresence experience and are relevant for 3GPP's IMS-based telepresence services from a media handling point of view. Furthermore, guidance is provided in [29] on the need for signalling these parameters at session initiation and during a session.

As far as possible, the audio/video parameters defined in [29] have been aligned with those specified by the IETF CLUE Working Group. However, there are several additional parameters that have also been specified. The set of parameters defined by ITU-T SG16 in [29] that are relevant for 3GPP's IMS-based telepresence services are documented below, and with also an indication their relationship to IETF CLUE.

5.3.2.1 Capture-Related Parameters

5.3.2.1.1 General parameters

Table 5.1: General parameters

Parameter	Need for signalling at session initiation	Need for signalling during session	Remarks
media Type	Y	Y	See the "MediaCapture" attributes in IETF CLUE data model schema [11].
captureScene description	Y	Y	See the "Description" attribute in 7.2.1 of IETF CLUE framework [8].
sceneView description	Y	Y	See the "View" attribute in 7.2.2 of IETF CLUE framework [8].
Lang	Y	N	See "Language" in IETF CLUE framework [8].
Priority	Y	Y	See "Priority" in IETF CLUE framework [8].
Embeddedtext	Y	Y	See "Embedded Text" in IETF CLUE framework [8].
relatedTo	Y	Y	See "Related to" in IETF CLUE framework [8].
Presentation	Y	Y	See "Presentation" in IETF CLUE framework [8].
personInfo	Y	Y	As per clause 7.1.1.10 in IETF CLUE framework [8].
personType	Y	Y	As per clause 7.1.1.11 in IETF CLUE framework [8].
sceneInformation	Y	Y	As per clause 7.3.1.1 in IETF CLUE framework [8].
mediaCapture description	Y	Y	See the "Description" attribute in clause 7.1.1 of IETF CLUE framework [8].
captureScene scale	Y	N	See "Scale Information" in IETF CLUE framework [8].
mediaCapture mobility	Y	N	See "Mobility of capture" in IETF CLUE framework [8].
mediaCapture view	Y	Y	See "View" in IETF CLUE framework [8].
maxGroupBandwidth	Y	N	See "maxGroupBandwidth" in IETF CLUE framework [8].
Simulcast	Y	Y	Telepresence systems may provide multiple encodings for the one capture through a technique known as simulcast. For example, this may be achieved by sending multiple video coding streams with different characteristics to allow a receiving endpoint to choose the stream that meets its needs. IETF CLUE WG has decided to utilize the mechanism defined in [38].

5.3.2.1.2 Visual parameters

Table 5.2: Visual parameters

Parameter	Need for signalling at session initiation	Need for signalling during session	Remarks
colorGamut	Y	N	This parameter indicates the Colour Gamut used in a Telepresence Video Stream. Signalled as part of the codec information, e.g. in H.264 and H.265 SEI.
lumaBitDepth	Y	N	This parameter indicates the bit depth of the luma samples in a digital picture. Signalled as part of the codec information, e.g. in H.264 and H.265 SEI.
chromaBitDepth	Y	N	This parameter indicates the bit depth of the chroma samples in a digital picture. Signalled as part of the codec information, e.g. in H.264 and H.265 SEI.
effectiveResolution	N	N	This parameter indicates effective resolution of a rendered video stream as perceived by the viewer. Not signalled.
captureArea	Y	Y	See "Area of Capture" in IETF CLUE framework [8].
capturePoint	Y	Y	See "Point of Capture" in IETF CLUE framework [8].
lineOfCapturePoint	Y	Y	See the "Point on line of Capture" attribute in IETF CLUE framework [8].
maxVideoBitrate	Y	Y	This parameter indicates the maximum number of bits per second relating to a single video encoding and is signalled in the SDP. See "max-mbps" in IETF RFC 6184 [30] and "CustomMaxMBPS" in ITU-T H.241 [31].
maxWidth	Y	N	This parameter indicates the maximum video resolution width in pixels and is signalled in the SDP. See "horizontal image size" in IETF RFC 6236 [32] and "CustomPictureFormat" in ITU-T H.245 [33].
maxHeight	Y	N	This parameter indicates the maximum video resolution height in pixels and is signalled in the SDP. See "vertical image size" in IETF RFC 6236 [32] and "CustomPictureFormat" in ITU-T H.245 [33].
maxFramerate	Y	N	This parameter indicates the maximum video framerate and is signalled in the SDP. See "framerate" in IETF RFC 4566 [34] and "MaxFPS" in ITU-T H.241 [31].

5.3.2.1.3 Audio parameters

Table 5.3: Audio parameters

Parameter	Need for signalling at session initiation	Need for signalling during session	Remarks
Audio capturePoint	Y	Y	See "Point of Capture" in IETF CLUE framework [8].
Audio lineOfCapturePoint	Y	Y	See the "Point on line of Capture" attribute in IETF CLUE framework [8].
Audio sensitivityPattern	Y	Y	See the "Audio Capture Sensitivity Pattern" attribute in IETF CLUE framework [8].
maxAudioBitrate	Y	Y	This parameter indicates the maximum number of bits per second relating to a single audio encoding and signalled in the SDP. See "bandwidth" in IETF RFC 4566 [34] and "maxBitRate" in ITU-T H.245 [33].
nominalAudio Level	TBD	TBD	This parameter indicates the nominal audio level sent in the Telepresence audio stream. See ITU-T H.245 [33].
dynamicAudioLevel	N	Y	This parameter indicates the actual audio level sent in the Telepresence audio stream as it varies as a function of time, and may be signalled in the RTP header extension. See IETF RFC 6464 [35].

5.3.2.1.4 Delay parameters

Table 5.4: Delay parameters

Parameter	Need for signalling at session initiation	Need for signalling during session	Remarks
endToEndVideoDelay	N	N	This parameter indicates the one-way end to end delay (camera lens to video display) of the video media sent between two Telepresence terminals. In order to provide a high QoE telepresence experience to end-users, telepresence systems, it is desirable for the end to end video delay to be less than 320 milliseconds. Not signalled.
endToEndAudioDelay	N	N	This parameter indicates the one-way end to end delay (mouth to ear) of the audio media sent between two Telepresence terminals. In order to provide a high QoE telepresence experience to end-users, telepresence systems, it is desirable for the end to end audio delay to be less than 280 milliseconds. Not signalled.
audioVideoSynchronization	N	N	This parameter indicates the synchronization between an audio and the corresponding video media stream (EndtoEndVideoDelay-EndtoEndAudioDelay). In order to provide high QoE telepresence services to end-users, telepresence systems should maintain synchronization within 40 and -60 milliseconds (i.e. synchronization error is less than 40 ms if the audio stream is ahead of the video stream and less than 60 ms if the video stream is ahead of the audio stream).Not signalled.

5.3.2.1.5 Multiple Source Capture parameters

Table 5.5: Multiple Source Capture parameters

Parameter	Need for signalling at session initiation	Need for signalling during session	Remarks
multContentCapture	Y	Y	See the ' Multiple content capture ' in IETF CLUE framework [8].
MCC sources	Y	Y	See the ' Multiple content capture ' in IETF CLUE framework [8].
MCC maxCaptures	Y	Y	See the ' Maximum Number of Captures within a MCC ' MCC attribute in IETF CLUE framework [8].
MCC policy	Y	Y	See the ' Policy ' MCC attribute in in IETF CLUE framework [8].
MCC synchronizationID	Y	Y	See the ' Synchronization Identity ' MCC attribute in IETF CLUE framework [8].

5.3.2.2 Telepresence System Environment parameters

Table 5.6: Telepresence System Environment parameters

Parameter	Need for signalling at session initiation	Need for signalling during session	Remarks
illuminantType	Y	Y	This parameter describes the profile of the visible light at a telepresence endpoint. May need to be signalled if lighting changes during session. Signalling is based on Annex E of ITU-T H.264 [36] and Annex E of ITU-T H.265 [37].
illuminantCRI Index	Y	Y	This parameter describes the colour rendering index (CRI) of the visible (ambient) light at the telepresence endpoint. Signalling is based on Annex E of ITU-T H.264 [36] and Annex E of ITU-T H.265 [37].
illuminantColourTemperature	Y	Y	This parameter describes the correlated colour temperature (CCT) of the visible (ambient) light at the telepresence endpoint. Signalling is based on Annex E of ITU-T H.264 [36] and Annex E of ITU-T H.265 [37].

5.3.3 Gap Analysis

The observed gaps from a media handling perspective include the following:

- Further requirements on TP UEs beyond those in TS 22.228 [3] may be defined and functional requirements from ITU-T SG16 on telepresence systems addressing user experience, statistical information reporting and network level aspects may be considered for this purpose.
- Audio/video parameters to be used by TP UEs in IMS-based telepresence sessions need to be profiled, considering IETF CLUE and ITU-T SG16 telepresence specifications including capture-related parameters and system environment related parameters.

5.4 MPEG Codecs Relevant for Telepresence

5.4.1 MPEG-4 AAC-ELD

5.4.1.1 Introduction

The Moving Picture Experts Group (MPEG) updated AAC-LD to AAC-ELD [20], providing more coding efficiency and lower algorithmic delay [40]. Interoperability is achieved through compliance to the MPEG-4 Audio Profile Low Delay AAC v2. When the clients are negotiating use of AAC-ELD, then SDP offers should include both AAC-ELD and AAC-LD, where AAC-ELD is preferred over AAC-LD.

5.4.1.2 Gap Analysis

Neither AAC-LD nor AAC-ELD is supported in MTSI. As AAC-ELD is an update of AAC-LD, which is supported in IR.39 (see clause 5.2.3 of the present document), and widely supported by mobile operating systems, AAC-ELD may also be considered for TP UEs.

5.4.2 MPEG Video Codecs

5.4.2.1 Introduction

IMS-based telepresence services are expected to support higher resolution video streams including High Definition (HD) video formats or even Ultra-High Definition (UHD) video formats, due to end terminals with typically larger screen sizes. Consequently, it is desirable to investigate the relevance of MPEG's advanced video codecs targeted for delivery of HD video and possibly UHD video streams over IMS-based telepresence services, with resolutions of 1080p or higher, and frame rates up to 60 fps.

In MTSI UEs, the current video codec requirements mandate support of H.264/AVC Constrained Baseline Profile (CBP), Level 1.2, and recommend support of H.264/AVC Constrained Baseline Profile (CBP), Level 3.1 and H.265/HEVC Main Profile, Main Tier, Level 3.1. In the context of IMS-based telepresence services, it is hence relevant to consider the following MPEG video codec profiles to be supported by TP UEs:

- H.264/AVC Constrained High Profile (CHP), Levels 1.3 or higher [36]
- H.265/HEVC Main Tier Main Profile, Levels 4.1 or higher [37]

5.4.2.2 Gap Analysis

It is possible that both mandatory and recommended codec requirements for TP UEs may be updated based on these additional MPEG video codec profiles.

6 Conclusion

Based on the gap analysis on media handling aspects of IMS-based telepresence, the following conclusions can be drawn toward normative specification work:

6.1 Codecs

Media codec requirements for IMS-based telepresence should be defined for speech, video and real-time text. Various gaps with respect to MTSI codec requirements were documented in the present document and these additional codec capabilities, e.g. as also documented in GSMA's IR.39 IMS Profile on High Definition Video Conference (see clause 5.2), should be considered during the course of the normative work. In particular, as discussed in clauses 5.2 and 5.4, it is relevant to consider the following codec profiles to be supported by TP UEs:

- Audio:
  - ITU-T G.719 [19] and MPEG AAC-LD [20] codecs from GSMA IR.39 profile [16]
  - MPEG AAC-ELD codec in [20]
- Video:
  - H.264/AVC Constrained High Profile (CHP), Levels 1.3 or higher [36]
  - H.265/HEVC Main Tier Main Profile, Levels 4.1 or higher [37]

6.2 Media Handling Aspects

Various gaps with respect to media handling aspects of telepresence were documented in the present document. Accordingly, the following areas should be within the scope of the normative work:

- Data transport aspects of a TP UE, including RTP/RTCP level requirements, user plane protocol stack aspects beyond the MTSI client, as identified in clause 5.1

- Audio/video parameters and their usage for IMS-based telepresence, as identified in clause 5.3
- Guidelines and examples for SDP and CLUE usage for IMS-based telepresence, e.g. similar to the guidelines and examples on SDP usage for MTSI in TS 26.114, as identified in clause 5.1

Annex A:
Change history

Change history							
Date	**TSG #**	**TSG Doc.**	**CR**	**Rev**	**Subject/Comment**	**Old**	**New**
2015-06	68	SP-150206			Presented at TSG SA#68 for information		1.0.0
2015-09	69	SP-150442			Presented at TSG SA#69 for approval	1.0.0	2.0.0
2015-09	69				Version 13.0.0	2.0.0	13.0.0
2015-12	70	SP-150645	0001	1	Corrections and Editorial Enhancements	13.0.0	13.1.0

Change history							
Date	**Meeting**	**TDoc**	**CR**	**Rev**	**Cat**	**Subject/Comment**	**New version**
2017-03	75					Version for Release 14	14.0.0
2018-06	80					Version for Release 15	15.0.0
2020-07	-	-	-	-	-	Update to Rel-16 version (MCC)	**16.0.0**
2022-04	-	-	-	-	-	Update to Rel-17 version (MCC)	**17.0.0**
2024-03	-	-	-	-	-	Update to Rel-18 version (MCC)	**18.0.0**

History

Document history		
V18.0.0	May 2024	Publication