**3GPP TSG- Meeting # *S4-211141***

**Online, -**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *CR-Form-v12.1* | | | | | | | | |
| **Draft CHANGE REQUEST** | | | | | | | | |
|  | | | | | | | | |
|  |  | **CR** |  | **rev** |  | **Current version:** |  |  |
|  | | | | | | | | |
| *For* [***HE******LP***](http://www.3gpp.org/3G_Specs/CRs.htm#_blank)*on using this form: comprehensive instructions can be found at* [*http://www.3gpp.org/Change-Requests*](http://www.3gpp.org/Change-Requests)*.* | | | | | | | | |
|  | | | | | | | | |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Proposed change affects:*** | UICC apps |  | ME | **X** | Radio Access Network |  | Core Network | **X** |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | | | | | |
| ***Title:*** | Corrections on ITT4RT | | | | | | | | | |
|  |  | | | | | | | | | |
| ***Source to WG:*** |  | | | | | | | | | |
| ***Source to TSG:*** |  | | | | | | | | | |
|  |  | | | | | | | | | |
| ***Work item code:*** | ITT4RT | | | | |  | ***Date:*** | | |  |
|  |  | | | |  | |  | | |  |
| ***Category:*** | **F** |  | | | | | ***Release:*** | | |  |
|  | *Use one of the following categories:* ***F*** *(correction)* ***A*** *(mirror corresponding to a change in an earlier release)* ***B*** *(addition of feature),* ***C*** *(functional modification of feature)* ***D*** *(editorial modification)*  Detailed explanations of the above categories can be found in 3GPP [TR 21.900](http://www.3gpp.org/ftp/Specs/html-info/21900.htm). | | | | | | | | *Use one of the following releases: Rel-8 (Release 8) Rel-9 (Release 9) Rel-10 (Release 10) Rel-11 (Release 11) … Rel-15 (Release 15) Rel-16 (Release 16) Rel-17 (Release 17) Rel-18 (Release 18)* | |
|  |  | | | | | | | | | |
| ***Reason for change:*** | | Some ABNF syntaxes are not matched with the asssociated normative text.. | | | | | | | | |
|  | |  | | | | | | | | |
| ***Summary of change:*** | | Some ABNF syntaxes are updated. It also include some editorial improvement. | | | | | | | | |
|  | |  | | | | | | | | |
| ***Consequences if not approved:*** | | The implementers find difficulty in the understanding of ITT4RT functionalities | | | | | | | | |
|  | |  | | | | | | | | |
| ***Clauses affected:*** | | 3.1, Y.2, Y.3, Y.6.1, Y.6.2.1, Y.6.2.2, Y.6.2.4, Y.6.2.5 | | | | | | | | |
|  | |  | | | | | | | | |
|  | | **Y** | **N** |  | | | |  | | |
| ***Other specs*** | |  | **X** | Other core specifications | | | | TS/TR ... CR ... | | |
| ***affected:*** | |  | **X** | Test specifications | | | | TS/TR ... CR ... | | |
| ***(show related CRs)*** | |  | **X** | O&M Specifications | | | | TS/TR ... CR ... | | |
|  | |  | | | | | | | | |
| ***Other comments:*** | |  | | | | | | | | |
|  | |  | | | | | | | | |
| ***This CR's revision history:*** | |  | | | | | | | | |

## \*\*\* Start change 1 \*\*\*

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:

NOTE: A term defined in the present document takes precedence over the definition of the same term, if any, in 3GPP TR 21.905 [1].

**example:** text used to clarify abstract rules by applying them literally.

**360-degree video:** A real-world visual scene captured by a set of cameras or a camera device with multiple lenses and sensors covering the sphere in all directions around the centre point of the camera set or camera device. The term 360-degree video may be used to include also limited 360-degree video.

**Limited 360-degree video:** A 360-degree video in which the visual scene does not cover the entire sphere around the center point of the camera set or camera device but only a part of it. A limited 360-degree video may be limited i) in the horizontal field to less than 360 degrees, or ii) in the vertical field to less than 180 degrees or iii) in both the vertical and horizontal fields.

**AMR, AMR-NB:** Both names refer to the AMR codec (3GPP TS 26.071 [11]) and are used interchangeably in this specification.

**Bitstream:** A bitstream that conforms to a video or audio encoding format.

**bitstream**: A sequence of bits that forms the representation of one or more coded video or audio sequences.

**CHEM:** The Coverage and Handoff Enhancements using Multimedia error robustness feature.

**Codec mode:** Used for the AMR and AMR-WB codecs to identify one specific bitrate. For example AMR includes 8 codec modes (excluding SID), each of different bitrate.

**Constrained terminal:** UE that is (i) operating in radio access capability category series "M" capable of supporting conversational services, and/or (ii) a wearable device which is constrained in size, weight or power consumption (e.g. connected watches), excluding smartphones and feature phones.

**DCMTSI client:** A data channel capable MTSI client supporting data channel media as defined in clause 6.2.10.

**DCMTSI client in terminal:** A DCMTSI client that is implemented in a terminal or UE. The term "DCMTSI client in terminal" is used in this document when entities such as MRFP, MRFC or media gateways are excluded.

**Dual-mono:** A variant of 2-channel stereo encoding where two instances of a mono codec are used to encode a 2-channel stereo signal.

**Evolved UTRAN:** Evolved UTRAN is an evolution of the 3G UMTS radio-access network towards a high-data-rate, low-latency and packet-optimized radio-access network.

**EVS codec:** The EVS codec includes two operational modes: EVS Primary operational mode (‘EVS Primary mode’) and EVS AMR-WB Inter-Operable (‘EVS AMR-WB IO mode’). When using EVS AMR-WB IO mode the speech frames are bitstream interoperable with the AMR-WB codec [18]. Frames generated by an EVS AMR-WB IO mode encoder can be decoded by an AMR-WB decoder, without the need for transcoding. Likewise, frames generated by an AMR-WB encoder can be decoded by an EVS AMR-WB IO mode decoder, without the need for transcoding.

**EVS Primary mode:** Includes 11 bit-rates for fixed-rate or multi-rate operation; 1 average bit-rate for variable bit-rate operation; and 1 bit-rate for SID (3GPP TS 26.441 [121]). The EVS Primary can encode narrowband, wideband, super-wideband and fullband signals. None of these bit-rates are interoperable with the AMR-WB codec.

**EVS AMR-WB IO mode:** Includes 9 codec modes and SID. All are bitstream interoperable with the AMR-WB codec (3GPP TS 26.171 ‎‎[17]).

**Field of View**: The extent of visible area expressed with vertical and horizontal angles, in degrees in the 3GPP 3DOF reference system as defined in TS 26.118 [180].

**Fisheye Video**: Video captured by a wide-angle camera lens that usually captures an approximately hemispherical field of view and projects it as a circular image.

**Frame Loss Rate (FLR):** The percentage of speech frames not delivered to the decoder. FLR includes speech frames that are not received in time to be used for decoding.

**ITT4RT client:** MTSI client supporting the Immersive Teleconferencing and Telepresence for Remote Terminals (ITT4RT) feature, as defined in Annex Y.

**ITT4RT-Tx client:** ITT4RT client only capable of sending immersive video.

**ITT4RT-Rx client:** ITT4RT client only capable of receiving immersive video

**ITT4RT MRF:** An ITT4RT client implemented by functionality included in the MRFC and the MRFP.

**ITT4RT client in terminal:** An ITT4RT client that is implemented in a terminal or UE. The term "ITT4RT client in terminal" is used in this document when entities such as ITT4RT MRF is excluded.

**Mode-set:** Used for the AMR and AMR-WB codecs to identify the codec modes that can be used in a session. A mode-set can include one or more codec modes.

**MSMTSI client:** A multi-stream capable MTSI client supporting multiple streams as defined in Annex S. An MTSI client may support multiple streams, even of the same media type, without being an MSMTSI client. Such an MTSI client may, for example, add a second video to an ongoing video telephony session as shown in Annex A.11. In that case, the MTSI client is an MSMTSI client only if it is fully compliant with Annex S.

**MSMTSI MRF:** An MSMTSI client implemented by functionality included in the MRFC and the MRFP.

**MSMTSI client in terminal:** An MSMTSI client that is implemented in a terminal or UE. The term "MSMTSI client in terminal" is used in this document when entities such as MRFP, MRFC or media gateways are excluded.

**MTSI client:** A function in a terminal or in a network entity (e.g. a MRFP) that supports MTSI.

**MTSI client in terminal:** An MTSI client that is implemented in a terminal or UE. The term "MTSI client in terminal" is used in this document when entities such as MRFP, MRFC or media gateways are excluded.

**MTSI media gateway (or MTSI MGW):** A media gateway that provides interworking between an MTSI client and a non MTSI client, e.g. a CS UE. The term MTSI media gateway is used in a broad sense, as it is outside the scope of the current specification to make the distinction whether certain functionality should be implemented in the MGW or in the MGCF.

**Omnidirectional media:** Media such as image or video and its associated audio that enable rendering according to the user's viewing orientation, if consumed with a head-mounted device, or according to user's desired viewport, otherwise, as if the user was in the spot where and when the media was captured.

**Operational mode:** Used for the EVS codec to distinguish between EVS Primary mode and EVS AMR-WB IO mode.

**Overlay:** A piece of visual media, rendered over omnidirectional video or image, or a viewport.

**Pose:** Position and rotation information associated to a viewport.

**Projected picture:** Picture that has a representation format specified by an omnidirectional video projection format.

**Projection:** Inverse of the process by which the samples of a projected picture are mapped to a set of positions identified by a set of azimuth and elevation coordinates on a unit sphere.

**Simulcast:** Simultaneously sending different encoded representations (simulcast formats) of a single media source (e.g. originating from a single microphone or camera) in different simulcast streams.

**Simulcast format:** The encoded format used by a single simulcast stream, typically represented by an SDP format and all SDP attributes that apply to that particular SDP format, indicated in RTP by the RTP header payload type field.

**Simulcast stream:** The RTP stream carrying a single simulcast format in a simulcast.

**Viewport**: Region of omnidirectional image or video suitable for display and viewing by the user.

## \*\*\* End change 1 \*\*\*

## \*\*\* Start change 2 \*\*\*

# Y.2 Architecture and Interfaces

Definitions, reference and coordinate systems, video signal representation and audio signal representation as described in clause 4.1 of TS 26.118 [180] are applicable.

Figure Y.1 provides a possible sender architecture that produces the RTP streams containing 360-degree video and immersive speech/audio as applicable to an ITT4RT client in terminal. VR content acquisition includes capture of 360-degree video and immersive speech/audio, as well as other relevant content such as overlays. Following VR content pre-processing and encoding of 360-degree video and immersive speech/audio components, the corresponding elementary streams are generated. For 360-degree projected video, pre-processing may include video stitching, rotation or other translations, and the pre-processed 360-degree video is then passed into the projection functionality in order to map 360-degree video into 2D textures using a mathematically specified projection format. Optionally, the resulting projected video may be further mapped region-wise onto a packed video. For 360-degree fisheye video, circular videos captured by fisheye lenses are not stitched, but directly mapped onto a 2D texture, without the use of the projection and region-wise packing functionalities (as described in clause 4.3 of ISO/IEC 23090-2 [179]). In this case, pre-processing may include arranging the circular images captured by fisheye lenses onto 2D textures, and the functionality for projection and mapping is not needed. For audio, no stitching process is needed, since the captured signals are inherently immersive and omnidirectional. Followed by the HEVC/AVC encoding of the 2D textures and EVS encoding of immersive speech/audio along with the relevant immersive media metadata (e.g., SEI messages), the consequent video and audio elementary streams are encapsulated into respective RTP streams and transmitted.

RTP stream for 360-Degree Video

Video pre-processing (e.g., stitiching, rotation, etc.)

Elementary stream (ES)

EVS Encoder

Acquisition

Video ES (including metadata)

RTP Encapsulation

Audio pre-processing

RTP Encapsulation

HEVC or AVC Encoder

RTP stream for Immersive Speech/Audio

Projection & Mapping

Audio ES

Figure Y.1 - Reference sender architecture for ITT4RT client in terminal

Figure Y.2 provides an overview of a possible receiver architecture that reconstructs the 360-degree video and immersive speech/audio in an ITT4RT client in terminal. Note that this figure does not represent an actual implementation, but a logical set of receiver functions. Based on one or more received RTP media streams, the UE parses, possibly decrypts and feeds the elementary video stream into the HEVC/AVC decoder and the speech/audio stream into the EVS decoder. The HEVC/AVC decoder obtains the decoder output signal, referred to as the "2D texture", as well as the decoder metadata. Likewise, the EVS decoder output signal contains the immersive speech/audio. The decoder metadata for video contains the Supplemental Enhancement Information (SEI) messages, i.e., information carried in the omnidirectional video specific SEI messages, to be used in the rendering phase. In particular, the decoder metadata may be used by the Texture-to-Sphere Mapping function to generate a 360-degree video (or part thereof) based on the decoded output signal, i.e., the texture. The viewport is then generated from the 360-degree video signal (or part thereof) by taking into account the pose information from sensors, display characteristics as well as possibly other metadata.

For 360-degree video, the following components are applicable:

- The RTP stream contains an HEVC or an AVC bitstream with omnidirectional video specific SEI messages. In particular, the omnidirectional video specific SEI messages as defined in ISO/IEC 23008-2 [119] and ISO/IEC 14496-10 [24] may be present.

- The video elementary stream(s) are encoded following the requirements in clause Y.3

Audio ES

RTP streams

RTP Receiver

Video ES (with metadata)

Texture to Sphere Mapping

Viewport Rendering for Immersive Video

EVS Decoder

Decoder output signal

Immersive Voice/audio Rendering

360-Degree Video

Figure Y.2 - Reference receiver architecture for ITT4RT- client in terminal

The output signal, i.e., the decoded picture or "texture", is then rendered using the Decoder Metadata information in relevant SEI messages contained in the video elementary streams as well as the relevant information signalled at the RTP/RTCP level (in the viewport-dependent case). The Decoder Metadata is used when performing rendering operations such as region-wise unpacking, projection de-mapping and rotation for 360-degree projected video, or fisheye video information for 360-degree fisheye video) toward creating spherical content for each eye. Details of such sample location remapping process operations are described in clause D.3.41.7 of ISO/IEC 23008-2 [119].

Viewport-dependent 360-degree video processing could be supported for both point-to-point conversational sessions and multiparty conferencing scenarios and can be achieved by sending from the ITT4RT-Rx client RTCP feedback messages with viewport information and then encoding and sending the corresponding viewport by the ITT4RT-Tx client or by the ITT4RT-MRF. This is expected to deliver resolutions higher than the viewport independent approach for the desired viewport. The transmitted RTP stream from the ITT4RT-Tx client or ITT4RT-MRF may also include the information on the region of the 360-degree video encoded in higher quality as the video generated, encoded and streamed by the ITT4RT-Tx client may cover a larger area than the desired viewport. Viewport-dependent processing is realized via RTP/RTCP based protocols that are supported by ITT4RT clients. The use of RTP/RTCP based protocols for viewport-dependent processing is further described in clause Y.7.2.NOTE: RTP header extensions for ITT4RT clients are FFS.

## \*\*\* End change 2 \*\*\*

## \*\*\* Start change 3 \*\*\*

# Y.3 Immersive 360-Degree Video Support

ITT4RT-Rx clients in terminals offering video communication shall support decoding capabilities based on:

- H.264 (AVC) [24] Constrained High Profile, Level 5.1 with the following additional restrictions and requirements on the bitstream:

- the maximum VCL Bit Rate is constrained to be 120 Mbps with cpbBrVclFactor and cpbBrNalFactor being fixed to be 1250 and 1500, respectively.

- the bitstream does not contain more than 10 slices per picture.

- H.265 (HEVC) [119] Main 10 Profile, Main Tier, Level 5.1.

In addition, ITT4RT-Rx clients in terminals may support:

- H.265 (HEVC) [119] Screen-Extended Main 10 Profile, Main Tier, Level 5.1.

- H.265 (HEVC) [119] Screen-Extended Main 4:4:4 10 Profile, Main Tier, Level 5.1.

ITT4RT-Tx clients in terminals offering video communication shall support encoding up to the maximum capabilities (e.g., color bit-depth, luma samples per second, luma picture size, frames per second) compatible with decoders compliant with the following on the bitstream:

- H.264 (AVC) [24] Constrained High Profile, Level 5.1 with the following additional restrictions and requirements:

- the maximum VCL Bit Rate is constrained to be 120 Mbps with cpbBrVclFactor and cpbBrNalFactor being fixed to be 1250 and 1500, respectively.

- the bitstream does not contain more than 10 slices per picture.

- H.265 (HEVC) [119] Main 10 Profile, Main Tier, Level 5.1.

In addition, ITT4RT-Tx clients in terminals may support:

- H.265 (HEVC) [119] Screen-Extended Main 10 Profile, Main Tier, Level 5.1.

- H.265 (HEVC) [119] Screen-Extended Main 4:4:4 10 Profile, Main Tier, Level 5.1.

Hence, for a Bitstream conforming to the H.264 (AVC) [24] Constrained High Profile, Level 5.1 delivered from an ITT4RT-Tx client to the ITT4RT-Rx client, the following restrictions apply:

- The profile\_idc shall be set to 100 indicating the High profile.

- The constraint\_set0\_flag, constraint\_set1\_flag, constraint\_set2\_flag and constraint\_set3\_flag shall all be set to 0, and constraint\_set4\_flag and constraint\_set5\_flag shall be set to 1.

- The value of level\_idc shall not be greater than 51 (corresponding to the level 5.1) and should indicate the lowest level to which the Bitstream conforms.

Furthermore, for a Bitstream conforming to the H.265 (HEVC) [119] Main 10 Profile, Main Tier, Level 5.1 delivered from an ITT4RT-Tx client to the ITT4RT-Rx client, the following restrictions apply:

- The general\_profile\_idc shall be set to 2 indicating the Main10 profile.

- The general\_tier\_flag shall be set to 0 indicating the Main tier.

- The value of level\_idc shall not be greater than 153 (corresponding to the Level 5.1) and should indicate the lowest level to which the Bitstream conforms.

For 360-degree video delivery across ITT4RT clients, the following components are applicable:

- The RTP stream shall contain an HEVC or an AVC bitstream with possible presence of omnidirectional video specific SEI messages. In particular, the omnidirectional video specific SEI messages as defined in clause D.2.41 of ISO/IEC 23008-2 [119] or ISO/IEC 14496-10 [24] may be present for the respective HEVC or AVC bitstreams.

- The video elementary stream(s) shall be encoded following the requirements in the Omnidirectional Media Format (OMAF) specification ISO/IEC 23090-2 [179], clauses 10.1.2.2 (viewport-independent case) or 10.1.3.2 (viewport-dependent case) for HEVC bitstreams and clause 10.1.4.2 for AVC bitstreams. Furthermore, the general video codec requirements for AVC and HEVC in clause 5.2.2 of TS 26.114 also apply.

ITT4RT-Rx clients are expected to be able to process the VR metadata carried in SEI messages for rendering 360-degree video according to the relevant processes. Relevant SEI messages contained in the elementary stream(s) with decoder rendering metadata may include the following information for the relevant processes as per clause D.3.41 of ISO/IEC 23008-2 [119] and ISO/IEC 14496-10 [24]:

- Projection mapping information (indicating the projection format in use, e.g., Equirectangular projection (ERP) or Cubemap projection (CMP)), for the projection sample location remapping process as specified in clauses 7.5.1.3 and 5.2 of ISO/IEC 23090-2 [179]

- Region-wise packing information (carrying region-wise packing format indication, any coverage restrictions or padding/guard region information in ithe packed picture), for the inverse processes of the region-wise packing as specified in clauses 7.5.1.2 and 5.4 of ISO/IEC 23090-2 [179]

- Sphere rotation information (indicating the amount of sphere rotation, if any, applied to the sphere signal before projection and region-wise packing at the encoder side), for the coordinate axes conversion process as specified in clause 5.3 of ISO/IEC 23090-2 [179]

- Frame packing arrangement (indicating the frame packing format for stereoscopic content), for the processes as specified in D.3.16 of ISO/IEC 23008-2 [119]

- Fisheye video information (indicating that the picture is a fisheye video picture containing a number of active areas captured by fisheye camera lens), for the fisheye sample location remapping process as specified in clause D.3.41.7.5 of ISO/IEC 23008-2 [119]

The exchange of SEI messages carrying VR metadata for rendering 360-degree video or fisheye video shall be performed using bitstream-level signalling as follows.

SEI messages shall be present in the respective video elementary streams corresponding to the HEVC or AVC bitstreams carrying 360-degree video or fisheye video from the ITT4RT-Tx client to the ITT4RT-Rx client, as per ISO/IEC 23008-2 [119] or ISO/IEC 14496-10 [24]. As expressed more clearly below, the mandatory inclusion of the specific SEI messages in the bitstream by the ITT4RT-Tx client and their decoder and rendering processing by the ITT4RT-Rx client is conditional upon successful SDP-based negotiation of the corresponding 360-degree video or fisheye video capabilities.

NOTE 1: The feasibility of the following SDP solution is FFS. “SEI messages may also be signalled in the SDP using the ‘sprop-sei’ parameter based on the procedures specified in IETF RFC 7798 [120] or via other SDP-based means, in the corresponding SDP offer or answer from the ITT4RT-Tx client to the ITTRT-Rx client, during the session setup involving media negotiations for 360-degree video or fisheye video. It should be noted that the signalling based on the ‘sprop-sei’ parameter is only available for the HEVC-based RTP payload formats and is not supported in AVC-based RTP payload formats as defined in IETF RFC 6184 [25].” More broadly, the feasibility of an out-of-band solution for signaling SEI message information is FFS. Such a solution may consider SDP-based approaches as well as those based on RTP header extensions.

NOTE 2: The frequency of the signalling of each SEI message is FFS.

In particular, the ITT4RT-Tx client supporting 360-degree video for viewport-independent processing shall signal in the bitstream the equirectangular projection SEI message (payloadType equal to 150) to the ITT4RT-Rx client, with the erp\_guard\_band\_flag set to 0.

If viewport-dependent processing (VDP) capability is successfully negotiated by the ITT4RT-Tx client and ITT4RT-Rx client for the exchange of 360-degree video, then, the ITT4RT-Tx client shall signal in the bitstream to the ITT4RT-Rx client either:

- the equirectangular projection SEI message (payloadType equal to 150) with the erp\_guard\_band\_flag set to 0, or

- the cubemap projection SEI message (payloadType equal to 151).

In order to optimize the spatial resolution of specific viewports, the ITT4RT-Tx client and ITT4RT-Rx client may negotiate the use of region-wise packing as part of the exchange of 360-degree video. If this is the case, the region-wise packing SEI message (payloadType equal to 155) shall also be signalled by the ITT4RT-Tx client to the ITT4RT-Rx client in the bitstream.

If stereoscopic video capability is successfully negotiated by the ITT4RT-Tx client and ITT4RT-Rx client as part of the exchange of 360-degree video, then the frame packing arrangement SEI message (payloadType equal to 45) shall also be signalled by the ITT4RT-Tx client to the ITT4RT-Rx client in the bitstream, with the following restrictions:

- The value of frame\_packing\_arrangement\_cancel\_flag is equal to 0.

- The value of frame\_packing\_arrangement\_type is equal to 4.

- The value of quincunx\_sampling\_flag is equal to 0.

- The value of spatial\_flipping\_flag is equal to 0.

- The value of field\_views\_flag is equal to 0.

- The value of frame0\_grid\_position\_x is equal to 0.

- The value of frame0\_grid\_position\_y is equal to 0.

- The value of frame1\_grid\_position\_x is equal to 0.

- The value of frame1\_grid\_position\_y is equal to 0.

Furthermore, ITT4RT-Tx clients supporting 360-degree fisheye video shall signal the fisheye video information SEI message (payloadType equal to 152) to the ITT4RT-Rx clients in the bitstream.

The bitstream delivered from an ITT4RT-Tx client to the ITT4RT-Rx client shall contain the corresponding SEI message and ITT4RT-Rx client shall process the VR metadata carried in the signalled SEI message(s) for rendering 360-degree video (provided the successful SDP-based negotiation of the corresponding 360-degree video or fisheye video capabilities associated with the SEI messages).

## \*\*\* End change 3 \*\*\*

## \*\*\* Start change 4 \*\*\*

Y.6.1 General

Based on the architecture described in clause Y.2, an SDP framework for immersive video and immersive voice/audio exchange for ITT4RT is presented to negotiate codec support, SEI messages for decoder rendering metadata, as well as RTP/RTCP signaling necessary for viewport dependent processing.

The SDP attributes *3gpp\_360video*, *3gpp\_fisheye,* *3gpp\_overlay*, *3gpp\_360bg* shall be used to indicate respectively a 360-degree projected video stream, a 360-degree fisheye video stream, a spherical overlay, and a 360-degree background image (or series of images or video). ITT4RT-Tx clients that support both 360-degree projected video and 360-degree fisheye video may include both *3gpp\_360video* and *3gpp\_fisheye* attributes as alternatives in the SDP offer, but an ITT4RT-Rx client willing to receive 360-degree video shall include only one attribute (either *3gpp\_360video* or *3gpp\_fisheye*, based on support or selection) in the SDP answer. *3gpp\_overlay* and *3gpp\_360bg* attributes may be included in the SDP answer independent on whether projected or fisheye video is selected, since spherical overlays and 360-degree background images are applicable to both types of 360-degree video streams. The detailed definition and usage of these SDP attributes are presented in the clauses below.

## \*\*\* End change 4 \*\*\*

## \*\*\* Start change 5 \*\*\*

Y.6.2.1 General

A new SDP attribute 3gpp\_360video is defined with the following ABNF syntax:

3gpp\_360video = "a=3gpp\_360video:" [SP FOV] [SP FOV\_CENTER] [SP "Stereo"] [VDP]

VDP = "VDP" [SP Projection] [SP PPM] SP viewport\_ctrl SP viewport\_size

NOTE: If the SDP negotiations become too complex, defining profiles can be considered.

The semantics of the above attribute and parameters is provided below. Unsupported parameters of the 3gpp\_360video attribute may be ignored.

An ITT4RT client supporting the 3gpp\_360video attribute shall support the following procedures:

- when sending an SDP offer, the ITT4RT client includes the 3gpp\_360video attribute in the media description for video in the SDP offer,

- when sending an SDP answer, the ITT4RT client includes the 3gpp\_360video attribute in the media description for video in the SDP answer if the 3gpp\_360video attribute was received in an SDP offer,

- after successful negotiation of the 3gpp\_360video attribute in the SDP, for the video streams based on the HEVC or AVC codec, the ITT4RT clients exchange an RTP-based video stream containing an HEVC or AVC bitstream with omnidirectional video specific SEI messages as defined in clause Y.3.

An ITT4RT client supporting the 3gpp\_360video attribute supporting use of viewport-dependent processing (VDP) shall include the VDP parameter in the SDP offer and answer. Depending on the value indicated by the VDP parameter, the ITT4RT client shall further support the following procedures:

- the RTCP feedback (FB) message described in clause Y.7.2 of type ‘Viewport’ to carry requested viewport information during the RTP streaming of media (signalled from the ITT4RT-Rx client to the ITT4RT-Tx client).

An ITT4RT client shall not include VDP parameter in the SDP answer if the SDP offer contains the 3gpp\_360video attribute without the VDP parameter.

An ITT4RT-Tx client that supports VDP may use viewport margins to maintain consistent quality during small head motion and also to reduce the need for frequent viewport updates. Viewport margins can be extended on all or some sides of the viewport and may be at the same quality (or resolution) as the viewport or at a quality (or resolution) lower than the viewport but higher than the background. Viewport margins may be extended around the viewport evenly or unevenly depending on head motion or network quality.

## \*\*\* End change 5 \*\*\*

## \*\*\* Start change 6 \*\*\*

Y.6.2.2 Projection

An ITT4RT- client supporting the 3gpp\_360video attribute with VDP supporting projection may include the Projection parameter indicating the types of projection (e.g. ERP, CMP) it prefers (in the order of preference) in the SDP. An ITT4RT client may respond to an SDP offer with multiple options indicated in the Projection parameter with the agreed option. An ITT4RT-Tx client is not required to provide the preferred form of projection indicated by an ITT4RT-Rx client but may do so when possible.

The ABNF syntax is defined as follows:

Projection = "projection=" proj-type \*("," proj-type)

proj-type = "ERP" / "CMP"

## \*\*\* End change 6 \*\*\*

## \*\*\* Start change 7 \*\*\*

Y.6.2.4 Picture Packing

An ITT4RT-Tx client may support sending the 360-degree video using tiled encoding, e.g., HEVC motion-constrained tiling. An ITT4RT-Tx client using the tiled encoding and supporting the 3gpp\_360video attribute shall include in its SDP offer the parameter PPM such that:

- A PPM value of 1 indicates mixed-quality tiled encoding is used

- A PPM value of 2 indicates mixed-resolution tiled encoding is used

Tiled encoding may be used to deliver the full 360-degree video or a high-quality video which includes the viewport and may include viewport margins.

An ITT4RT-Tx client sending low-quality background 360-degree video frame-packed with a high-quality viewport (possibly with margins) such that the two regions have overlapping content shall include in its SDP offer the parameter PPM. In this case the PPM is an ordered comma-seperated list of the following six fields:

- PPWHQ defines packed\_picture\_width of the high-quality region in pixels

- PPHHQ defines packed\_picture\_height of the high-quality region in pixels

- TRHQ defines transform operations applied on the high-quality region.

- PPWLQ defines packed\_picture\_width of the low-quality region in pixels

- PPHLQ defines packed\_picture\_height of the low-quality region in pixels

- TRLQ defines transform operations applied on the low-quality region

The transform operations have a value of 0-7 as defined in Table Y.6.1.

Table Y.6.1: Transform values

|  |  |
| --- | --- |
| **Transform value** | **Transform operation** |
| 0 | no transorm |
| 1 | mirrored horizontally |
| 2 | rotation by 180 degrees (counter-clockwise) |
| 3 | rotation by 180 degrees (counter-clockwise) before mirroring horizontally |
| 4 | rotation by 90 degrees (counter-clockwise) before mirroring horizontally |
| 5 | rotation by 90 degrees (counter-clockwise) |
| 6 | rotation by 270 degrees (counter-clockwise) before mirroring horizontally |
| 7 | rotation by 270 degrees (counter-clockwise) |

An ITT4RT-Rx client shall render the high-quality viewport region where these two regions are overlapping. The PPM parameter for defining the HQ and LQ regions should be used when the information remains constant during the session. When the packed regions are not overlapping, the high-quality and low-quality regions do not need to be explicitly defined and SEI messages for region-wise packing may be used instead of the SDP PPM parameter.

The ABNF syntax is defined as follows:

PPM = "ppm=" ppm-value

ppm-value = "1" / "2" / transforms

transforms = "[" PPWHQ "," PPHHQ "," TRHQ "," PPWLQ "," PPHLQ "," TRLQ "]"

PPWHQ = pos-integer

PPHHQ = pos-integer

TRHQ = transform-value

PPWLQ = pos-integer

PPHLQ = pos-integer

TRLQ = transform-value

pos-integer = POS-DIGIT \*DIGIT

POS-DIGIT = %x31-39 ;1-9

transform-value = "0" / "1" / "2" / "3" / "4" / "5" / "6" / "7"

; transfrom values as per Table Y.6.1

Note: The size of the viewport and fov attributes define the size of projected regions. It should be considered if the two values should be included explicitly here

## \*\*\* End change 7 \*\*\*

## \*\*\* Start change 8 \*\*\*

Y.6.2.5 Viewport Control

An ITT4RT client that supports the 3gpp\_360video with VDP shall in its SDP offer include the parameter viewport\_ctrlwith one or more of the following control options:

- *device\_controlled* if ITT4RT-Tx client will provide VDP based on the requested viewport indicated by the RTCP feedback (FB) message type ‘Viewport’ sent by the corresponding ITT4RT-Rx client.

- *recommended\_viewport* if ITT4RT-Tx client will provide VDP with the help of a recommendation/prediction engine.

- *presenter\_viewport* if ITT4RT\_Tx will provide VDP based on the viewport of an ITT4RT-Rx client other than the one the SDP offer is being sent to.

Table Y.6.2 provides a mapping between viewport control values and viewport control options..

Table Y.6.2: Viewport control values

|  |  |
| --- | --- |
| **Viewport control value** | **Viewport control option** |
| 0 | device\_controlled |
| 1 | recommended\_viewport |
| 2 | presenter\_viewport |

Multiple options are provided as a comma-separated list. An ITT4RT client that receives an SDP offer with multiple viewport\_ctrl options may include its preferred viewport\_ctrl option in the SDP answer. If no options are given in the answer, the sender shall use the first option in the list. If the recommended\_viewport is successfully negotiated as viewport\_ctrl, the ITT4RT-Rx client should not use viewport prediction when sending the RTCP feedback (FB) message type ‘Viewport’ to avoid any conflicts with the prediction engine of the ITT4RT-Tx client. The ABNF syntax is defined as follows:

viewport\_ctrl = "viewport\_ctrl=" vc-value \*2 ("," [SP] vc-value)

vc-value = "0" / "1" / "2"

; viewport control values as per Table Y.6.2

## \*\*\* End change 8 \*\*\*

## \*\*\* End of changes \*\*\*