3GPP TSG-SA4 Meeting #110e *S4-201198*

19-28 August 2020 revision of S4-201196

**Agenda item** 11.5

**Source:** Intel

**Title:** ITT4RT: Proposed Conclusions for Permanent Document

**Document for** Discussion and Agreement

# Introduction

ITT4RT permanent document (PD) contains requirements, working assumptions and potential solutions on various features around real-time 360-degree video delivery protocols and formats.

During Rel-16 timeframe, ITT4RT PD was launched following the official start of the work item in Jan. 2019, with the understanding that normative work shall start during Rel-17. So far, all contributions addressing work item objectives were handled without any specific structure or priority towards keeping the work scope as flexible as possible. This approach has resulted in many new and exciting functionality now documented as part of the latest version of the PD in S4-201024.

Now that we are in Rel-17 timeframe and should begin normative specification development with actual CRs to the impacted specifications (i.e., TS 26.114 and TS 26.223), we believe there is a need for a more structured approach to managing the work progress and accordingly handling input contributions under a specific prioritization framework. In this input document, we offer some perspectives in this direction toward starting the discussion among the contributing companies and agreeing on a way forward. It is expected that the work item timeplan will also be revised accordingly in order to reflect the agreed-upon priorities.

# Proposed Way Forward

We propose a phased approach toward normatively introducing ITT4RT functionalities into the impacted specifications (i.e., TS 26.114 and TS 26.223) via CRs as follows:

**Phase 1 (Target completion date: By end of SA4#112)**

1- Develop a basic real-time 360-degree video and limited support for immersive voice/audio delivery framework in TS 26.114 with

1. Recommendations of audio and video codec configurations (e.g., profile, level, and encoding constraints of EVS, HEVC, AVC as applicable) to deliver high quality VR experiences
2. Constraints on media elementary streams and RTP encapsulation formats
3. Recommendations of SDP configurations for negotiating of immersive video and voice/audio capabilities.
4. Use of omnidirectional video specific Supplemental Enhancement Information (SEI) messages for carriage of metadata required for rendering of the omnidirectional video
5. An appropriate signalling mechanism, e.g., RTP/RTCP-based, for indication of viewport information to enable viewport-dependent media processing and delivery

New input into the permanent document will be considered during Phase 1, but will be de-prioritized against the CRs fulfilling the criteria above.

Relevant potential solutions for the immersive video have been documented in the permanent document as follows.

* Sections 6.1 and 8: Basic immersive video functionality (i.e., viewport-independent), SDP negotiation framework, elementary stream and RTP encapsulation formats, video codec configurations
* Sections 6.1, 9.1-9.6, 9.12: RTCP-based viewport information signaling and SDP negotiation framework for viewport-dependent processing
* Section 6.2: Signaling of camera calibration parameters
* Sections 6.3 and 6.4: SDP-based solution for signaling of overlays and still background

It is expected that these solutions will be the basis of the CRs on immersive video, but other technical solutions may also be considered toward (i) updating the permanent document, (ii) inclusion as part of CR(s).

Since the IVAS codec cannot be finalized in the Rel-17 timeframe, a solution should be specified providing only limited support for immersive voice/audio using the EVS codec based on multi-mono EVS coding (requires characterization), and also based on mono EVS coding as fallback (which is already fully characterized). Such a solution has not yet been documented in the PD, so this is an item requiring urgent progress. It is essential that this is done and normative specification in TS 26.114 is completed during Phase 1. If however, the solution for limited immersive voice/audio support cannot be completed during Phase 1, it shall be moved into Phase 2 as discussed below. In this case, the agreements on the basic 360-degree video solution shall be kept in draft CR(s) during Phase 1, and official CR agreements will be postponed into Phase 2.

No further new use cases and requirements will be considered, i.e., use cases and requirements will essentially be technically frozen following SA4#111e. Only possible clarifications or minor technical updates to the existing use cases and requirements will be considered.

**Phase 2 (Target completion date: By end of SA4#114, but could be extended until end of Rel-17 if necessary)**

Develop a more enhanced real-time 360-degree video delivery framework in TS 26.114 based on the following potential solutions in the permanent document

* Section 9.7: Viewport margins
* Section 9.8: QoE metrics around motion to high quality delay and other relevant metrics reporting functionality
* Section 9.9: Alternative viewport information signaling using WebRTC data channel
* Section 9.10: Viewport sharing and following
* Section 10: Overlays using scene description
* Section 6.5: Network-based stitching through use of NBMP
* Any other new features adopted in the permanent document during Phase 1

It is expected that any CR(s) toward TS 26.223 will also be considered during this phase.

If the limited immersive voice/audio support as described above cannot be finalized during Phase 1, this work shall be a high priority item for Phase 2 to be completed by the end of Rel-17. At a minimum, limited immersive voice/audio support shall be based on mono EVS coding, which is already characterized. However, the work target is to enable limited immersive speech/audio support based on multi-mono EVS, and to also carry out the associated quality / performance characterization for dual-mono EVS.

It is the goal that official CR agreements will take place starting at the time when both a basic 360-degree solution and limited immersive voice/audio support can be introduced into TS 26.114 and TS 26.223. When IVAS is complete, the expectation is that full support for immersive voice/audio will be added, but this is likely to happen in a later release as part of a different work item.

No new input into the permanent document will be considered during Phase 2. In the meantime, the PD content may be further refined with editorial improvements and technical clarifications and ideally published as a new 3GPP technical report (like it was done for MMCMH before).