3GPP TSG SA WG-4 # 122 Meeting S4- S4-230075

20th – 24th February 2023 Revision of S4aV230004

Source: Dolby Laboratories Inc., Qualcomm Incorporated, AT&T, Philips International B.V., VoiceAge Corporation, Nokia Corporation, Xiaomi

Title: New WID on Split Rendering for Immersive Audio

Document for: Agreement

Agenda Item: 6.2

3GPP™ Work Item Description

Information on Work Items can be found at <http://www.3gpp.org/Work-Items>   
See also the [3GPP Working Procedures](http://www.3gpp.org/specifications-groups/working-procedures), article 39 and the TSG Working Methods in [3GPP TR 21.900](http://www.3gpp.org/ftp/Specs/html-info/21900.htm)

Title: Split Rendering for Immersive Audio

Acronym: SRIA

Unique identifier:

{A number to be provided by MCC at the plenary}

Potential target Release: *Rel-18*

# 1 Impacts

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Affects: | UICC apps | ME | AN | CN | Others (specify) |
| Yes |  | X |  | X |  |
| No | X |  |  |  |  |
| Don't know |  |  | X |  |  |

# 2 Classification of the Work Item and linked work items

## 2.1 Primary classification

### This work item is a …

|  |  |
| --- | --- |
| X | Feature |
|  | Building Block |
|  | Work Task |
|  | Study Item |

## 2.2 Parent Work Item

For a brand-new topic, use “N/A” in the table below. Otherwise indicate the parent Work Item.

|  |  |  |  |
| --- | --- | --- | --- |
| Parent Work / Study Items | | | |
| Acronym | Working Group | Unique ID | Title (as in 3GPP Work Plan) |
| MeCAR | SA4 | 950015 | Media Capabilities for Augmented Reality |
| IVAS\_Codec | SA4 | 770024 | EVS Codec Extension for Immersive Voice and Audio Services |

### 2.3 Other related Work Items and dependencies

|  |  |  |
| --- | --- | --- |
| Other related Work /Study Items (if any) | | |
| Unique ID | Title | Nature of relationship |
| 880011 | Study on 5G Glass-type AR/MR Devices | Study on the support of AR/MR with 5G glass-type devices. TR 26.998 concludes 5G Real-time Communication as an area for potential standardisation. |
| 960045 | Split Rendering Media Service Enabler | SR\_MSE is considered orthogonal as it explicitly excludes transcoding or AR conversational use cases. No specific impact is expected on SR\_MSE work or vice-versa. |
| 830005 | Terminal Audio quality performance and Test methods for Immersive Audio Services | No specific impact is expected on ATIAS work or vice-versa. Acoustic performance and test methods of/for glass-type devices would fully remain under ATIAS scope. |

Dependency on non-3GPP (draft) specification:

None

# 3 Justification

Work currently carried out under the MeCAR and related work items assumes a common XR Baseline Client architecture. An essential characteristic is that for both visual and audio media components a functional split is envisioned between a Presentation Engine comprising a set of composite renderers (“pre-renderer”) that are controlled by a Scene Manager and an XR Runtime performing a set of functions that interface with a platform to perform commonly required post-rendering operations (“post-renderer”) prior to final output. Part of the XR Runtime is an Audio Subsystem that may in its post-rendering operations adjust the audio data received from the Presentation Engine to match the visual rendering. The relevant interface between Presentation engine and XR Runtime may be a 5G physical interface between, e.g., between a smartphone or 5G EDGE and AR glasses like those considered in 5G EDGe-Dependent AR (EDGAR) and 5G Wireless Tethered AR UEs as described in 3GPP TR 26.998.

Diagram

Description automatically generated

Figure - XR Baseline Client architecture [S4-221577]

The functional split is a result of stringent implementation and operational requirements applicable for rendering of XR media on AR glasses. On the one hand, the operations on the glasses are strictly complexity constrained. On the other hand, the interface between Presentation Engine and XR Runtime may impose other constraints. For instance, the MeCAR WID highlights a latency from the scene renderer (Presentation Engine) to device rendering (XR Runtime) in the range of 50-60ms and sensor and device data streaming to the network in order to support network-based processing of device sensor information (i.e., pre-rendering relying on head-tracker information received upstream). It is notable that the transmission over the interface may generally be bit rate constrained and dependent on the specific physical interface.

Audio rendering comprises of signal processing functionalities that include:

* Binauralization of audio input based on head rotation (3DoF),
* Binauralization of audio input based on listener position and head rotation (6DoF),
* Room acoustics synthesis.

Audio input to be rendered may be a combination of diegetic immersive (3D audio) and non-diegetic sounds. The diegetic immersive sounds need to be binauralized using the up-to-date head rotation data. It is essential to avoid head rotation data delay. The head rotation data is typically originating from the head-tracker available from the lightweight device operating XR Runtime. The room acoustic synthesis can be performed using room impulse response data or parametric representation thereof, typically supplied to the Presentation Engine.

Given the strict constraints, the lightweight devices would typically not be capable of performing all the binauralization processing and room acoustic synthesis.

Depending on constraints and design preferences of the lightweight device (AR glasses, earbuds, etc.) and the properties of the interface between Presentation Engine and XR Runtime, solutions are needed for split rendering of immersive audio meeting the following example requirements:

* Provision of a pre-rendering component executed in the Presentation Engine accepting upstream received control data (head-tracker data, room acoustics data) to render the immersive audio streams into (a) suitable intermediate representation(s),
* Provision of a light-weight post-rendering component executed in the XR Runtime to convert the intermediate representation(s) into a binaural representation matching the latest head-tracker data available in the glasses,
* Capability to provide a low motion to sound latency, as defined in 3GPP TR 26.918,
* High audio quality,
* Scalable bit rate support to deal with different possible physical interfaces.

Another aspect is the currently ongoing standardization of the EVS Codec Extension for Immersive Voice and Audio Services (IVAS) codec. While split rendering is not a specific design objective, the IVAS codec work item should ideally provide split rendering solutions that would enable using IVAS services over clients with split architecture.

Bearing in mind the evolution of the AR/XR technologies, it would be desirable to design split renderer solutions that under certain limitations are agnostic in a sense that the pre-render component could be connected with any immersive binaural audio framework through suitable APIs.

Thus, the work to be carried out under this work item should develop split render solutions for immersive audio that give benefits to devices and services with split renderer architecture. In that sense, the work item will generally contribute to the goals of its parent work item MeCAR. However, no upwards dependency on MeCAR is expected, neither in terms of impact on timelines nor on objectives. Timewise, it is expected that results of the work item may be brought to already completed MeCAR specifications by means of CRs.

# 4 Objective

The overall objective of this work item is to develop solutions for immersive audio split rendering that are compatible with the envisaged split architectures. The following objectives should be achieved with the work item:

* Provide format specification for intermediate split rendered audio representation(s).
  + Provide functional requirements for pre-renderer operations to be carried out by Presentation Engine.
  + Define suitable APIs.
* Provide encoder, bitstream and decoder specification for intermediate split rendered representations including audio and post-rendering control metadata.
* Provide a post-renderer specification for decoded split renderer representations to provide binaural audio output in response to head-tracker input and post-rendering control metadata.
* Specify the interfacing/integration with IVAS codec. Note that this may require adding constraints to IVAS codec work like, e.g., adding a deliverable of a split renderer solution compatible with the objectives of this work item or characterizing aspects of the solution under the IVAS codec work.

The work item shall in a first phase identify and agree relevant requirements to be documented in a TR. This shall cover:

* Design constraints related to complexity and memory as well as constraints related to relevant 5G interfaces between Presentation Engine and XR Runtime such as bit rate, latency, down- and upstream traffic characteristics.
* Design constraints related to functional capability requirements such as rendering of non-diegetic sounds, 3DoF rendering of diegetic immersive sounds, 6DoF rendering of diegetic immersive sounds, including simultaneous rendering of different sound categories.
* Performance requirements.

As part of the first phase, IVAS (external) renderer and external renderer interface for split rendering are characterized. Subsequently (a) solution(s) meeting the requirements shall be specified with demonstrated applicability to IVAS with integration of the post-renderer in IVAS (external) renderer, within the timelines of the IVAS codec work item. In a second phase, further solutions should be developed to address potential gaps and/or enhance the solution(s) developed in the first phase.

The developments under this work item shall lead to a new specification defining among others textual descriptions of the involved renderers and codec (incl. frame loss concealment) of the intermediate split rendered audio representation(s). The requirements will be documented in a first technical report. The performance of the developed solutions in relation to the requirements will be documented in a second technical report. Parts of the solutions and performance evaluations results related to phase 1 work may be referenced through IVAS codec specifications/technical report. The developed solutions should also be referenced in the MeCAR specification.

# 5 Expected Output and Time scale

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| New specifications | | | | | |
| Type | TS/TR number | Title | For info  at TSG# | For approval at TSG# | Rapporteur |
| TR | 26.xxx | Split Rendering for Immersive Audio; Requirements |  | SA#101 (June 2023) | TBD |
| TS | 26.xxx | Split Rendering for Immersive Audio | SA#102 (Dec 2023)  Including phase 1 solutions | SA#103 (Mar 2024)  Including phase 1+phase 2 solutions | TBD |
| TR | 26.xxx | Split Rendering for Immersive Audio; Performance characterization | SA#102 (Dec 2023)  Including phase 1 solutions results | SA#103 (Mar 2024)  Including phase 1+phase 2 solutions results | TBD |

|  |  |  |  |
| --- | --- | --- | --- |
| Impacted existing TS/TR | | | |
| TS/TR No. | Description of change | Target completion plenary# | Remarks |
| TS 26.119 | Referencing of Split Rendering Solutions for Immersive Audio | SA#103 (Mar 2024) |  |

# 6 Work item Rapporteur(s)

TBD

# 7 Work item leadership

SA4

# 8 Aspects that involve other WGs

None

# 9 Supporting Individual Members

|  |
| --- |
| Supporting IM name |
| Dolby Laboratories Inc. |
| Qualcomm Incorporated |
| AT&T |
| Philips International B.V. |
| VoiceAge Corporation |
| Nokia Corporation |
| Xiaomi |