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: Rendering of Immersive Audio for AR

Acronym: RIAAR

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 a functional split is envisioned between a Presentation Engine comprising a set of composite renderers 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 operations, e.g. post-rendering, prior to final output. 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 a lightweight device (AR glasses) like those considered in 5G EDGe-Dependent AR (EDGAR) and 5G Wireless Tethered AR UEs as described in 3GPP TR 26.998.

The functional split assumed in split renderer architectures is a result of stringent implementation and operational requirements applicable for rendering of XR media on AR glasses. For head-tracked immersive audio, the need to rely on a split renderer architecture, may depend on various factors among which the round-trip latency between the renderer in the presentation engine and the lightweight device is a decisive parameter. There are scenarios where this latency is substantial and exceeding acceptable motion to sound latency limits (see TR 26.918) in which the choice is to adopt a split rendering approach for audio in a similar way as for video or to carry out audio rendering on the lightweight device provided it does not exceed its strict complexity constraints. In other scenarios, that latency may be low, in which case head-tracked binaural pre-rendering can be done by the renderer in the presentation engine while the post-rendering by the lightweight device means directly outputting the pre-rendered binaural audio signal. It is notable that the transmission over the interface may generally be bit rate constrained and dependent on the specific physical interface.

Binaural audio rendering comprises of signal processing functionalities that may 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.

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 low complex 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 component executed in the XR Runtime to convert the intermediate representation(s) into a binaural representation sufficiently well matching the latest head-tracker data available in the glasses,
* Capability to provide a sufficiently low motion to sound latency, as defined in 3GPP TR 26.918, for a range of relevant round-trip latency scenarios between pre-renderer in presentation engine and the lightweight device.
* 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 low complex rendering for lightweight devices is not a specific design objective, the IVAS codec work item should ideally provide solutions that would enable using IVAS services over head-tracked lightweight clients meeting relevant requirements.

Bearing in mind the evolution of the AR/XR technologies, it would be desirable to design low complex solutions for head-tracked binaural audio rendering on lightweight devices that under certain limitations are agnostic in a sense that the pre-renderer component in the presentation engine could be connected with any immersive binaural audio framework through suitable APIs.

# 4 Objective

The overall objective of this work item is to develop solutions for low complex immersive binaural audio on head-tracked lightweight devices that are compatible with the envisaged split architectures and demonstrate operational benefits over solutions with full decoding and rendering in the XR Runtime. The following objectives should be achieved with the work item:

* Provide format specification for intermediate 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 representations including audio with and without post-rendering control metadata.
* Provide a post-renderer specification for decoded split renderer representations to provide binaural audio output with and without head-tracker input and post-rendering control metadata.

Consider potential solutions offered by the IVAS work item, if available, and specify the necessary interfaces.

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.

The solution(s) are characterized for the range of relevant interface characteristics between presentation engine and lightweight device. The case where the immersive audio is decoded and rendered within the XR Runtime should be considered as a reference.

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 representation(s). 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 may be referenced through, e.g., 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) | SA#103 (Mar 2024) | TBD |
| TR | 26.xxx | Split Rendering for Immersive Audio; Performance characterization | SA#102 (Dec 2023) | SA#103 (Mar 2024) | 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 |