|  |  |
| --- | --- |
| 3GPP TR ab.cde V0.2.0 (2023-03) | |
| Technical Report | |
| 3rd Generation Partnership Project;  Technical Specification Group SA;  Immersive Audio for Split Rendering Scenarios; Requirements (Release 18) | |
|  | |
|  |  |
|  | |
| The present document has been developed within the 3rd Generation Partnership Project (3GPP TM) and may be further elaborated for the purposes of 3GPP. The present document has not been subject to any approval process by the 3GPPOrganizational Partners and shall not be implemented. This Specification is provided for future development work within 3GPPonly. The Organizational Partners accept no liability for any use of this Specification. Specifications and Reports for implementation of the 3GPP TM system should be obtained via the 3GPP Organizational Partners' Publications Offices. | |

|  |
| --- |
|  |
| ***3GPP***  Postal address  3GPP support office address  650 Route des Lucioles - Sophia Antipolis  Valbonne - FRANCE  Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16  Internet  https://www.3gpp.org |
| ***Copyright Notification***  No part may be reproduced except as authorized by written permission. The copyright and the foregoing restriction extend to reproduction in all media.  © 2023, 3GPP Organizational Partners (ARIB, ATIS, CCSA, ETSI, TSDSI, TTA, TTC).  All rights reserved.  UMTS™ is a Trade Mark of ETSI registered for the benefit of its members  3GPP™ is a Trade Mark of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners LTE™ is a Trade Mark of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners  GSM® and the GSM logo are registered and owned by the GSM Association |

Contents

[Foreword 5](#_Toc130473931)

[Introduction 6](#_Toc130473932)

[1 Scope 7](#_Toc130473933)

[2 References 7](#_Toc130473934)

[3 Definitions of terms, symbols and abbreviations 7](#_Toc130473935)

[3.1 Terms 7](#_Toc130473936)

[3.2 Symbols 7](#_Toc130473937)

[3.3 Abbreviations 8](#_Toc130473938)

[4 Interfaces 8](#_Toc130473939)

[5 Physical Design Constraints 8](#_Toc130473940)

[6 Functional Design Constraints 8](#_Toc130473941)

[7 Performance Requirements 8](#_Toc130473942)

# Foreword

This Technical Report 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.

In the present document, modal verbs have the following meanings:

**shall** indicates a mandatory requirement to do something

**shall not** indicates an interdiction (prohibition) to do something

The constructions "shall" and "shall not" are confined to the context of normative provisions, and do not appear in Technical Reports.

The constructions "must" and "must not" are not used as substitutes for "shall" and "shall not". Their use is avoided insofar as possible, and they are not used in a normative context except in a direct citation from an external, referenced, non-3GPP document, or so as to maintain continuity of style when extending or modifying the provisions of such a referenced document.

**should** indicates a recommendation to do something

**should not** indicates a recommendation not to do something

**may** indicates permission to do something

**need not** indicates permission not to do something

The construction "may not" is ambiguous and is not used in normative elements. The unambiguous constructions "might not" or "shall not" are used instead, depending upon the meaning intended.

**can** indicates that something is possible

**cannot** indicates that something is impossible

The constructions "can" and "cannot" are not substitutes for "may" and "need not".

**will** indicates that something is certain or expected to happen as a result of action taken by an agency the behaviour of which is outside the scope of the present document

**will not** indicates that something is certain or expected not to happen as a result of action taken by an agency the behaviour of which is outside the scope of the present document

**might** indicates a likelihood that something will happen as a result of action taken by some agency the behaviour of which is outside the scope of the present document

**might not** indicates a likelihood that something will not happen as a result of action taken by some agency the behaviour of which is outside the scope of the present document

In addition:

**is** (or any other verb in the indicative mood) indicates a statement of fact

**is not** (or any other negative verb in the indicative mood) indicates a statement of fact

The constructions "is" and "is not" do not indicate requirements.

# Introduction

[

An essential architectural characteristic of XR clients is the reliance on a functional split between a set of composite pre-renderers that are implemented as parts of a presentation engine and a set of post-rendering operations implemented on a lightweight end-device (AR glasses) prior to final output.

The functional split assumed in split renderer architectures is a result of stringent implementation and operational requirements applicable for rendering of XR media on XR devices. 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 may be substantial which may prefer a split rendering approach with pose correction in the end device for binaural audio in a similar way as for video unless decoding and head-tracked binaural audio rendering on the lightweight device does not exceed its strict complexity constraints. In other scenarios, that latency may be sufficiently low, in which case the head-tracked binaural rendering can exclusively be done in the presentation engine. It is notable that the transmission over the interface may generally be bit rate constrained and dependent on the specific physical interface.

Technical solutions for the provision of immersive audio in split rendering scenarios are largely dependent on the associated properties and imposed requirements, which are identified by the present TR.

]

# 1 Scope

[

The present document identifies relevant requirements associated with split rendering scenarios for immersive audio. It covers:

* Design constraints related to complexity and memory as well as constraints related to relevant interfaces between presentation engine and end device 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, and room acoustics synthesis.
* Performance requirements.
* ]

# 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”.

…

[x] <doctype> <#>[ ([up to and including]{yyyy[-mm]|V<a[.b[.c]]>}[onwards])]: “<Title>”.

It is preferred that the reference to TR 21.905 be the first in the list.

# 3 Definitions of terms, symbols and abbreviations

This clause and its three (sub) clauses are mandatory. The contents shall be shown as “void” if the TS/TR does not define any terms, symbols, or abbreviations.

## 3.1 Terms

For the purposes of the present document, the terms given in 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 TR 21.905 [1].

Definition format (Normal)

**<defined term>:** <definition>.

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

## 3.2 Symbols

For the purposes of the present document, the following symbols apply:

Symbol format (EW)

<symbol> <Explanation>

## 3.3 Abbreviations

For the purposes of the present document, the abbreviations given in 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 TR 21.905 [1].

Abbreviation format (EW)

<ABBREVIATION> <Expansion>

# 4 Interfaces

## [4.1 Connectivity scenarios

### 4.1.1 Introduction

Immersive Audio for Split Rendering may be present in a variety of connectivity scenarios between the presentation engine and audio producing devices. This clause provides a non-exhaustive list of the envisioned connectivity scenarios.

Device design types, ordered by media capability performances, are currently:

* Device type 1: 5G AR glasses (standalone)
* Device type 2: 5G UE-tethered AR glasses with AR/MR application on AR Glass
* Device type 3: 5G UE-tethered AR glasses with AR/MR application on tethered device
* Device type 4: 5G XR HMD UE (standalone)

Associate Editor’s note: Those device design types are further described in S4-230738 (MeCAR Permanent Document v7.0). The Device type names are aligned with the agreed clause 10 in S4-230920 (TS 26.119 v0.2.0). It is proposed to add a direct reference to those documents for the Device Type descriptions or copy those descriptions here.

NOTE: With all device types, playback of the binaural acoustic signals can happen through built-in loudspeakers or through a separate, connected device (true wireless stereo earbuds, wireless headphones or wired headphones/earbuds). While specification and requirements on the proprietary link are outside the scope of 3GPP, it is expected that a properly implemented system would allow for a seamless transition between audio playback through built-in speakers and audio playback through earbuds/headphones.

### 4.1.2 Device Type 1 – 5G AR glasses (standalone)

Device type 1 refers to standalone AR glasses. Such devices may be able to support simpler audio decoding capabilities [TBD] and can feature head-tracking but may not be capable of complex video or audio processing.

The connectivity scenarios are the same as those of Device Type 4 (see Clause 4.1.4), except that not all scenarios may be possible to implement due to the physical constraints.

Associate Editor’s note: It is expected that some guidance as to what scenarios from Device Type 4 are recommended once the physical constraints are better understood.

Diagram

Description automatically generated

### 4.1.2 Device Type 2 – 5G UE-tethered AR glasses with AR/MR application on AR Glass

Device type 2 refers to an XR viewer that runs an AR/MR application and is tethered to a 5G device that provides connectivity and possibly additional processing. The tethering may be wireless or wired, but it is proprietary.

The audio connectivity scenarios are like those of Device Type 3 (see Clause 4.1.3), except that not all scenarios may be possible to implement due to the physical constraints.

Associate Editor’s note: It is expected that some guidance as to what scenarios from Device Type 3 are recommended once the physical constraints are better understood.

Waterfall chart

Description automatically generated

### 4.1.3 Device Type 3 - 5G UE-tethered AR glasses with AR/MR application on tethered device

The XR viewer is tethered to a 5G device (Remote Device) that includes the application and the XR functions. The tethering may be wireless or wired, but it is proprietary.

The 5G device runs the application that uses the Media access and rendering capabilities of the 5G device to run an AR/MR experience. The AR glass is connected to the 5G Device, but the XR runtime API is exposed to the 5G device/phone.

Diagram

Description automatically generated

#### 4.1.3.1 Scenario 1 – Local Decoding and Rendering on 5G Phone

*Associate Editor’s note: This is Scenario 4 in contribution S4-230842.*

The fourth scenario is as follows in Figure 2-4, refer to a type of “5G WireLess Tethered AR UE” in TR26.998[2], the Phone UE connects to Cloud/Edge through an embedded 5G modem, the Phone UE and Lightweight UE connect through WiFi or 5G sidelink, maybe through Bluetooth for audio, and the Phone UE provides the capabilities of both decoding and rendering.

A picture containing text, screenshot, font, black

Description automatically generated

Figure 1 - Figure 2-4 Phone-dependent Lightweight UE

**Variation A**

*Associate Editor’s note: Add text/figures in similar format as for Device Type 4. Audio playback through built-in speakers*

**Variation B**

*Associate Editor’s note: Add text/figure in similar format as for Device Type 4. Audio playback through earbuds.*

**Variation C**

*Associate Editor’s note: Add text/figure in similar format as for Device Type 4. Audio playback and pose estimation through earbuds.*

#### 4.1.3.2 Scenario 2 – Local Decoding and Pre-Rendering on 5G Phone

*Associate Editor’s note: This is Scenario 5 in contribution S4-230842.*

The fifth scenario is as follows in Figure 2-5, refer to a type of “5G WireLess Tethered AR UE” in TR26.998[2], the Phone UE connects to Cloud/Edge through an embedded 5G modem, the Phone UE and Lightweight UE connect through WiFi or 5G sidelink, maybe through Bluetooth for audio, Lightweight UE sends Pose Information to Phone UE if needed, and the Phone UE and Lightweight UE provide the capabilities of decoding and rendering together.

A picture containing text, screenshot, font

Description automatically generated

Figure 2 - Figure 2-5 Phone-dependent Lightweight UE

**Variation A**

*Associate Editor’s note: Add text/figures in similar format as for Device Type 4. Audio playback through built-in speakers*

**Variation B**

*Associate Editor’s note: Add text/figure in similar format as for Device Type 4. Audio playback through earbuds.*

**Variation C**

*Associate Editor’s note: Add text/figure in similar format as for Device Type 4. Audio playback and pose estimation through earbuds.*

#### 4.1.3.3 Scenario 3 – Decoding and Pre-Rendering on Cloud/Edge, 5G Phone as a pass-through device

*Associate Editor’s note: This is Scenario 6 in contribution S4-230842.*

The sixth scenario is as follows in Figure 2.6, refer to a type of “5G WireLess Tethered AR UE” in TR26.998[2], the Phone UE connects to Cloud/Edge through an embedded 5G modem, the Phone UE and Lightweight UE connect through WiFi or 5G sidelink, maybe through Bluetooth for audio, Lightweight UE sends Pose Information to Cloud/Edge if needed, and the Cloud/Edge and Lightweight UE provide the capabilities of decoding and rendering together, the Phone UE just acts as a relay device.

A picture containing text, screenshot, font

Description automatically generated

Figure 3 - Figure 2-6 Cloud/Edge-dependent Lightweight UE

**Variation A**

*Associate Editor’s note: Add text/figures in similar format as for Device Type 4. Audio playback through built-in speakers*

**Variation B**

*Associate Editor’s note: Add text/figure in similar format as for Device Type 4. Audio playback through earbuds.*

**Variation C**

*Associate Editor’s note: Add text/figure in similar format as for Device Type 4. Audio playback and pose estimation through earbuds.*

#### 4.1.3.4 Scenario 4 – Decoding and Pre-Rendering on Cloud/Edge, further decoding, and pre-rendering on 5G Phone

*Associate Editor’s note: This is Scenario 7 in contribution S4-230842. Is this a valid scenario? Seems to apply pre-rendering twice.*

The seventh scenario is as follows in Figure 2-7, refer to a type of “5G WireLess Tethered AR UE” in TR26.998[2], the Phone UE connects to Cloud/Edge through an embedded 5G modem, the Phone UE and Lightweight UE connect through WiFi or 5G sidelink, maybe through Bluetooth for audio, Lightweight UE sends Pose Information to Phone UE and Cloud/Edge if needed, and the Cloud/Edge, Phone UE and Lightweight UE provide the capabilities of decoding and rendering together.

A close-up of a phone

Description automatically generated with low confidence

Figure 4 - Figure 2-7 Cloud/Edge and Phone UE-dependent Lightweight UE

**Variation A**

*Associate Editor’s note: Add text/figures in similar format as for Device Type 4. Audio playback through built-in speakers*

**Variation B**

*Associate Editor’s note: Add text/figure in similar format as for Device Type 4. Audio playback through earbuds.*

**Variation C**

*Associate Editor’s note: Add text/figure in similar format as for Device Type 4. Audio playback and pose estimation through earbuds.*

#### 4.1.3.5 Scenario 5 – Decoding and Pre-Rendering on Cloud/Edge, rendering on 5G Phone

*Associate Editor’s note: This is Scenario 8 in contribution S4-230842. Is this a valid scenario? Seems to apply pre-rendering twice.*

The eighth scenario is as follows in Figure 2-8, refer to a type of “5G WireLess Tethered AR UE” in TR26.998[2], the Phone UE connects to Cloud/Edge through an embedded 5G modem, the Phone UE and Lightweight UE connect through WiFi or 5G sidelink, maybe through Bluetooth for audio, Lightweight UE sends Pose Information to Phone UE and Cloud/Edge if needed, and the Cloud/Edge and Phone UE provide the capabilities of decoding and rendering together.

A close-up of a phone

Description automatically generated with low confidence

Figure 5 - Figure 2-8 Cloud/Edge and Phone UE-dependent Lightweight UE

### 4.1.4 Device Type 4 – 5G XR HMD UE (Standalone)

For Device Type 4, the 5G device runs the XR application that uses the Media access and rendering capabilities of the 5G device to run an XR experience. All XR functionalities are included in a single device.



#### 4.1.4.1 Scenario 1 – Local Decoding and Rendering

The first scenario is as follows in Figure 2-1, (also referred as “5G Standalone AR UE” in TR26.998[2]). The 5G XR HMD UE connects directly to Cloud/Edge through an embedded 5G modem, and the 5G XR HMD UE has local capabilities of both decoding and rendering.

A picture containing text, screenshot, font, logo

Description automatically generated

**Variation A – Rendering, pose estimation and audio playback on HMD**

In Variation A, the pose estimate, audio decoding, binaural rendering and audio playback are all performed by the 5G XR HMD UE.



Figure 6: 5G XR HMD UE with local decoding/rendering

**Variation B – Decoding, Rendering and pose estimation on HMD, audio playback on earbuds.**

In Variation B, the Standalone XR HMD Device is connected to a pair of True Wireless Stereo (TWS) earbuds or headphones through means of a wireless or wired proprietary link. The TWS earbuds/headphones plays back the binaural audio instead of the built-in speakers used in Variation A, but the pose estimation function is still performed by the 5G XR HMD UE.



Figure 7: 5G XR HMD UE with local decoding/rendering and audio playback through TWS earbuds/headphones

NOTE: Variation B is expected to be more prevalent than Variation C described below due to overall lower motion to sound latency and possibly a better pose estimation capability by the 5G XR HMD UE.

**Variation C – Decoding and Rendering on HMD, pose estimation and audio playback on earbuds.**

In Variation C, the Standalone XR HMD Device is connected to a pair of TWS Earbuds or headphones through means of a wireless or wired proprietary link. The TWS earbuds/headphones performs pose estimation that is provided to the 5G XR HMD UE. The 5G XR HMD UE performs binaural rendering using that pose estimate and the earbuds/headphones playback the binaural audio.



Figure 8: 5G XR HMD UE with local decoding/rendering and audio playback / pose estimation through TWS earbuds/headphones

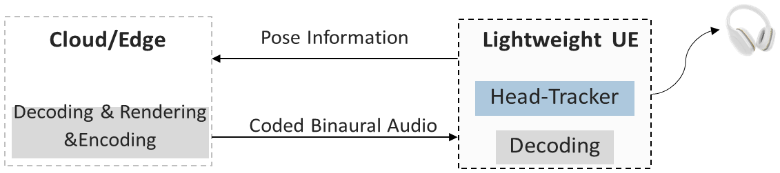
**Variation D – Decoding and Pre-Rendering on HMD, Rendering and audio playback on earbuds**

If the overall motion to sound latency is expected to be high for variations A, B and C, use of audio split rendering between the 5G XR HMD UE and the earbuds/headphones may be appropriate.

*Associate Editor’s note: Add figure for Variation D.*

#### 4.1.4.2 Scenario 2 – Decoding and Rendering on Cloud/Edge

The second scenario is as follows in Figure 2-2 and refers to a type of “5G EDGe-Dependent AR UE” in TR26.998[2]. The Standalone XR HMD Device connects directly to Cloud/Edge through an embedded 5G modem, and the Cloud/Edge provides the capabilities of both decoding and rendering.



**Variation A – Decoding and Rendering on Cloud/Edge, pose estimation and playback on HMD**

In Variation A, the pose estimate is obtained at the Standalone XR HMD UE and sent all the way to the cloud/edge which performs audio decoding, binaural rendering, and stereo encoding for transmission to the Standalone XR HMD UE. Potentially high motion to sound latency can happen in this case since it includes the time it takes for the pose estimate to be transmitted to the cloud/edge and the binaural audio to be transmitted to the UE.



Figure 9 - 5G XR HMD UE with Cloud/Edge based decoding and rendering

**Variation B - Decoding and Rendering on Cloud/Edge, pose estimation on HMD, playback on earbuds**

In Variation B, a pair of earbuds/headphones is used to playback the binaural audio instead of the built-in speakers used in Variation A, but the pose estimation function is still performed by the 5G XR HMD UE. Variation B is expected to be more prevalent than Variation C described below, due to lower motion to sound latency and possibly better pose estimation capability by the 5G XR HMD UE.

The 5G XR HMD UE may be a pass-through device for the audio.



Figure 10 - 5G XR HMD UE with Cloud/Edge based decoding and rendering and audio playback through TWS Earbuds/headphones

**Variation C - Decoding and Rendering on Cloud/Edge, pose estimation and playback on earbuds, HMD as a pass-through.**

In Variation C, the 5G XR HMD UE is connected to TWS Earbuds or headphones through means of a wireless or wired proprietary link. The earbuds/headphones can perform pose estimation and the Standalone XR HMD Device may act as a pass-through device for both the downlink stereo audio and the uplink pose estimation data.



Figure 11 - 5G XR HMD UE with Cloud/Edge based decoding and rendering and pose estimation and audio playback through TWS Earbuds/headphones

#### 4.1.4.2 Scenario 3 – Decoding and Pre-Rendering on Cloud/Edge

The third scenario is as follows in Figure 2-3, refer to a type of “5G EDGe-Dependent AR UE” in TR26.998[2], the Lightweight UE connects directly to Cloud/Edge through an embedded 5G modem, and the Cloud/Edge and Lightweight UE provide the capabilities of decoding and rendering together.

A screenshot of a computer

Description automatically generated with medium confidence

**Variation A - Decoding and Pre-Rendering on Cloud/Edge, pose estimation and playback on HMD**

In Variation A, the pose estimate is obtained at the 5G XR HMD UE and sent all the way to the cloud/edge which performs audio decoding, pre-rendering, and re-encoding (ISAR Encoder) for transmission to the 5G XR HMD UE. Motion to sound latency can be partially compensated, since the 5G XR HMD UE can provide pose correction and binaural rendering (ISAR Decoder). Audio playback is through the built-in loudspeakers.



Figure 12 - 5G XR HMD UE with Cloud/Edge based pre-rendering and local rendering/pose correction and audio playback

**Variation B - Decoding and Pre-Rendering on Cloud/Edge, pose estimation on HMD and playback on earbuds**

In Variation B, a pair of earbuds/headphones is used to playback the binaural audio instead of the built-in speakers used in Variation A, but the pose estimation function is still performed by the 5G XR HMD UE. Variation B is expected to be more prevalent than Variation C described below due to possibly better pose estimation capability by the 5G XR HMD UE.



Figure 13- 5G XR HMD UE with Cloud/Edge based pre-rendering, local rendering/pose correction and audio playback through TWS Earbuds/headphones

**Variation C - Decoding and Pre-Rendering on Cloud/Edge, pose estimation and playback on earbuds, HMD as a pass-through**

In Variation C, a pair of earbuds/headphones performs pose estimation and playback the binaural audio instead of the built-in speakers used in Variation A. The Standalone XR HMD acts as a passh-through device between the Cloud/Edge and the earbuds.



Figure 14 - 5G XR HMD UE with Cloud/Edge based pre-rendering and rendering, pose correction and audio playback through TWS Earbuds/headphones

## 4.2 Audio Architectures for Split Rendering Scenarios

### 4.2.1 Introduction

An XR scene usually comprises both visual and audio media. Within the scope of ISAR the visual media follows a split rendering approach, where decoding and (pre-)rendering are performed by a capable device (e.g., an edge server), and limited processing with lower complexity is performed on the lightweight UE.

For the immersive audio media different constraints in terms of complexity and memory as well as constraints related to relevant interfaces between remote presentation engine and end device such as bit rate, latency (including motion-to-sound latency), down- and upstream link characteristics may apply.

The following generic architectures illustrate the separation of decoding and rendering of the downstream audio between lightweight UE and capable devices, limited to the data flow relevant to the application of the pose information for head-tracked binaural audio.

NOTE: Pose information may be required for the media generation and thus be sent upstream to the core network (and any media encoding instances beyond) independent of the three architectures.

Selection of an architecture has an impact on complexity and memory as well as applicability to relevant interfaces between remote presentation engine and end device due to bit rate, latency (including motion-to-sound latency), down- and upstream traffic characteristics.

### 4.2.2 Local Audio Rendering

The immersive audio data is streamed directly to the lightweight UE, which is responsible for decoding, rendering, and synchronizing the audio with the corresponding visual content. The lightweight UE processes the pose information locally and adjusts the audio rendering accordingly to create a convincing immersive experience.

NOTE: This architecture is not a split architecture for the audio media in the sense that complex operations are offloaded to a capable device. This also represents the case that should be considered as the reference.

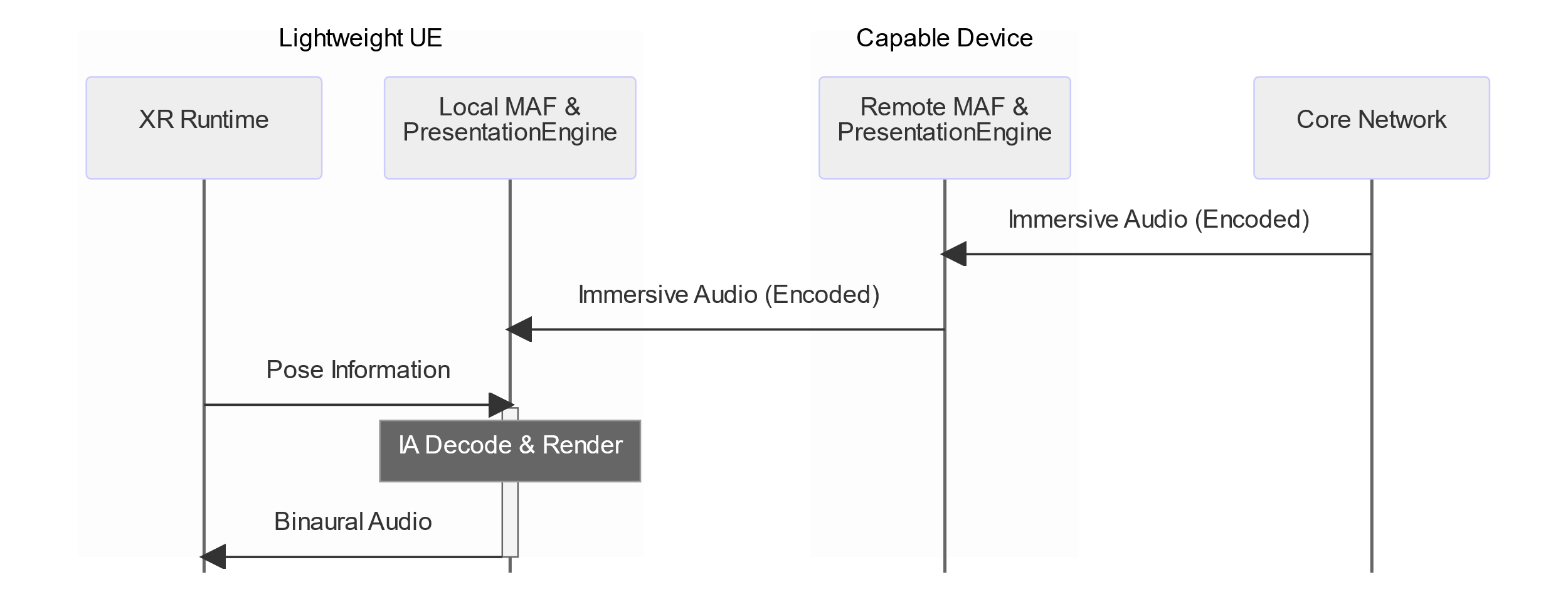


Figure 3.2-1: Sequence of data flow for Architecture 1, Local Audio Rendering

### 4.2.3 Distributed Audio Rendering

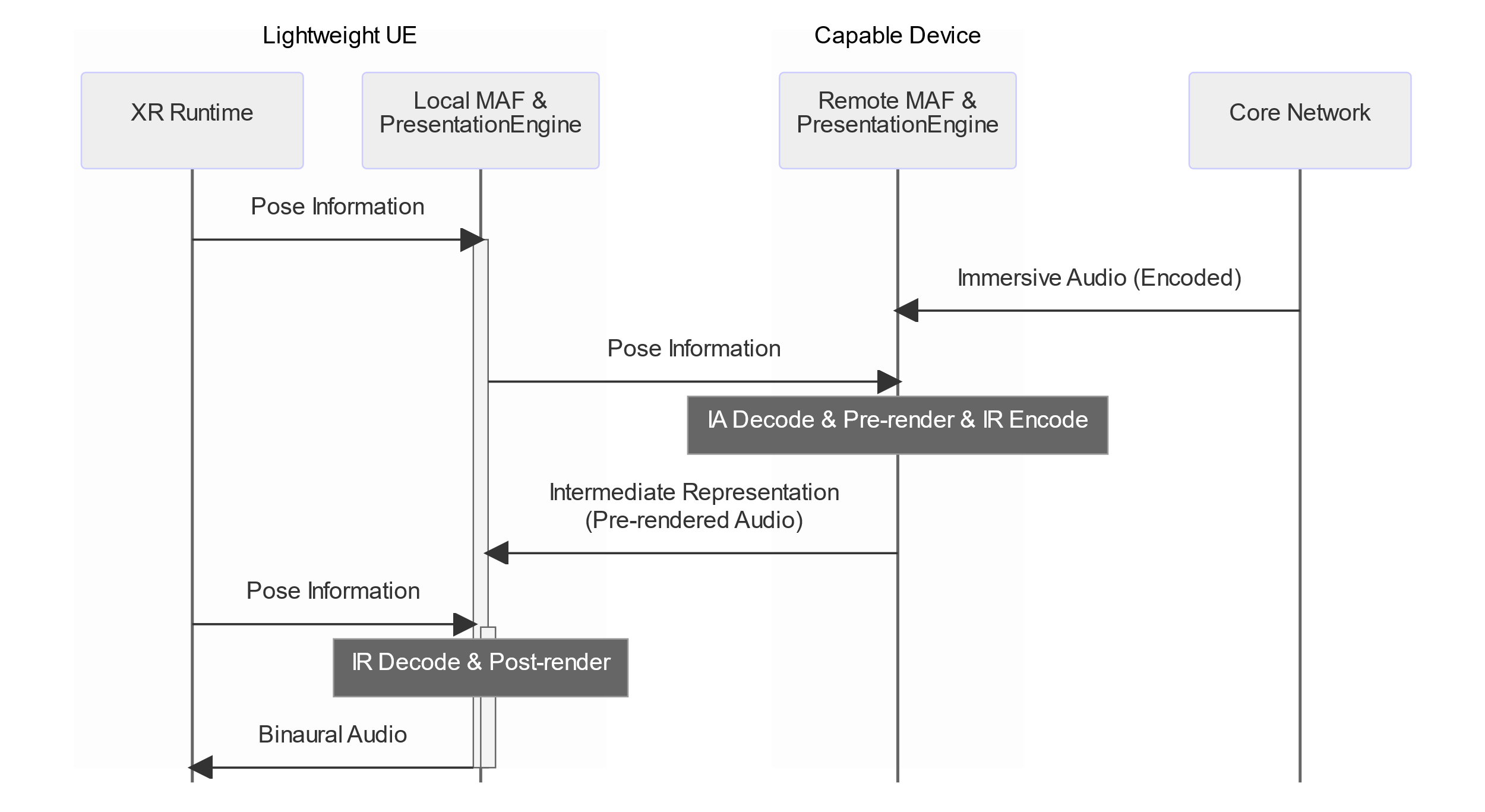
The capable device performs decoding and pre-rendering of the immersive audio media, and the pre-rendered audio is transmitted to the lightweight UE. The pose information is sent to the capable device if needed, which adjusts the pre-rendering based on the pose data to generate an ‘intermediate representation’. The lightweight UE then performs decoding of the received intermediate representation and applies post-rendering for pose correction using a recent pose information. 

Figure 3.3-1: Sequence of data flow for Architecture 2, Distributed Audio Rendering

### 4.2.4 Remote Audio Rendering

The capable device is responsible for decoding and fully rendering the immersive audio media and encoding the rendered audio into an ‘intermediate representation’, containing coded binaural audio. The intermediate representation is transmitted to the lightweight UE, which performs decoding of the rendered media. The lightweight UE synchronizes the binaural audio with the corresponding visual content.

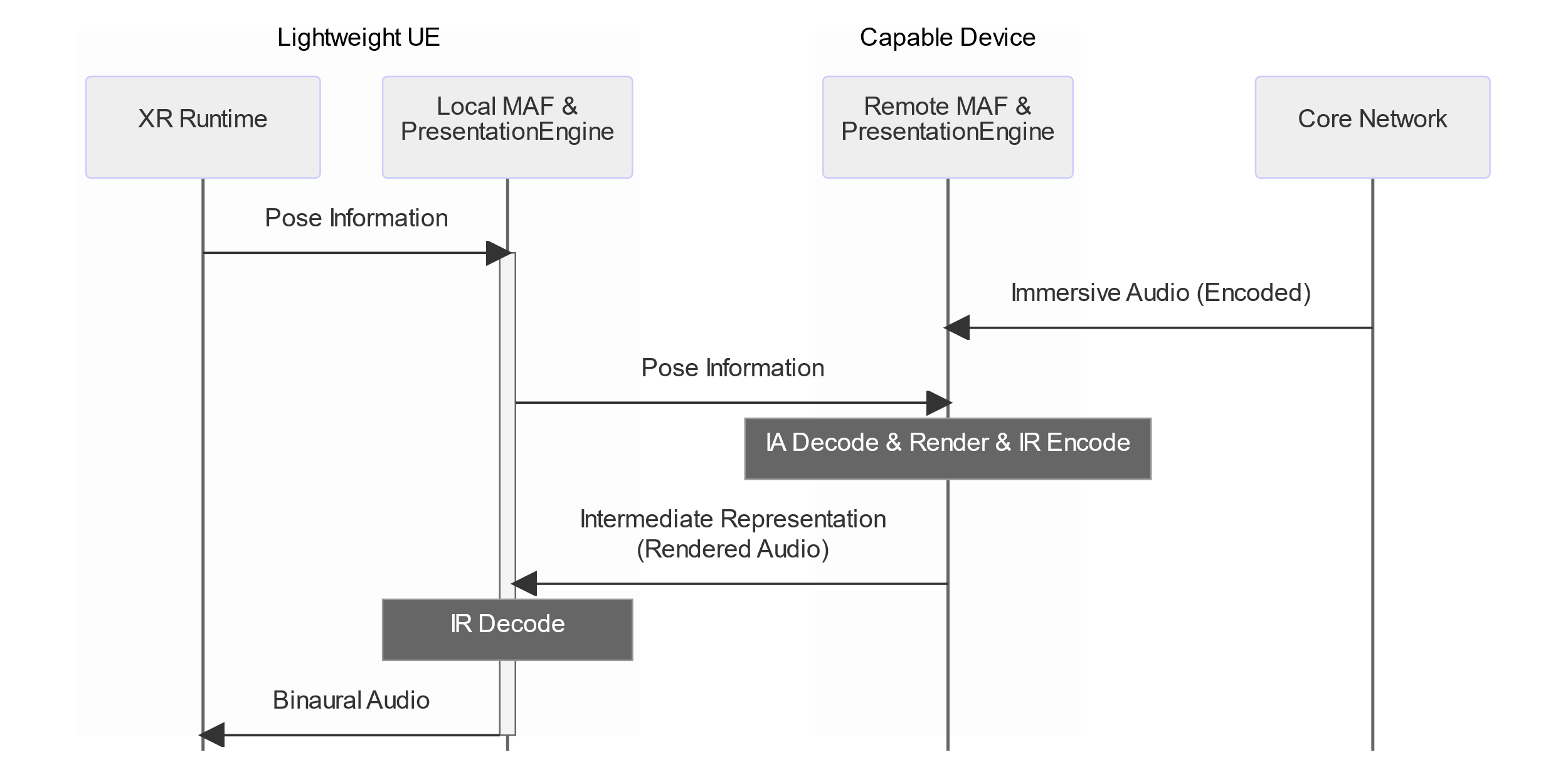


Figure 3.4-1: Sequence of data flow for Architecture 3, Remote Audio Rendering]

Editor’s Note:

* *Identify relevant interfaces between presentation engine and end device.*

# 5 Physical Design Constraints

Editor’s Note:

* *Identify design constraints related to complexity and memory*
* *Identify design constraints related to relevant interfaces such as bit rate, latency, down- and upstream traffic characteristics*
* *Identify implementation-specific design preferences*

# 6 Functional Design Constraints

Editor’s Note:

* *Identify functional design constraints 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, and room acoustics synthesis*

# 7 Performance Requirements

Editor’s Note:

* *Identify performance requirements such as subjective and objective quality targets*

Annex <A>:  
<Informative annex title for a Technical Report>

Informative annexes in Technical Reports do not use "(informative") in the title, since all annexes in TRs are informative. Use style "Heading 9" in TRs.

Annex <X>:  
Change history

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Change history | | | | | | | |
| Date | Meeting | TDoc | CR | Rev | Cat | Subject/Comment | New version |
| 2023-03 | SA4-e (AH) Audio SWG post 122 |  |  |  |  | Initial skeleton version | 0.0.1 |
| 2023-03 | SA4-e (AH) Audio SWG post 122 |  |  |  |  | Audio SWG agreed initial skeleton version | 0.0.2 |
| 2023-04 | SA4#123-e | S4-230 |  |  |  |  |  |