A logo of a company

AI-generated content may be incorrect. ISO/IEC JTC 1/SC 29/WG 6 N0357

**ISO/IEC JTC 1/SC 29/WG 6**

**MPEG Audio Coding Convenorship: DE**

**Document type:** Output Document

**Title:** Technologies under Consideration for ISO/IEC 14496-3:2019/CD AMD1

**Status:** Approved

**Date of document:** 2025-07-04

**Source:** ISO/IEC JTC 1/SC 29/WG 6

**Expected action:** None

**Action due date:** None

**No. of pages:** 12 (with cover page)

**Email of Convenor:** Thomas.Sporer@idmt.fraunhofer.de

**Committee URL:** <https://isotc.iso.org/livelink/livelink/open/jtc1sc29wg6>

**INTERNATIONAL ORGANISATION FOR STANDARDISATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC 1/SC 29/WG 6**

**MPEG AUDIO CODING**

**ISO/IEC JTC 1/SC 29/WG 6 N0357**

**July 2025, Daejeon, Korea**

|  |  |
| --- | --- |
| **Title** | **Technologies under Consideration for ISO/IEC 14496-3:2019/CD AMD1** |
| **Source** | **WG 6, MPEG Audio Coding** |
| **Status** | **Approved** |
| **Serial Number** | **25415** |

# Summary

This document describes the technologies that have been proposed for inclusion into ISO/IEC 14496-3 AMD1 at the 20th WG6 meeting. In addition, it contains a high-level summary of what interested experts intend to be working on until the 21st WG6 meeting, as well as a proposed draft amendment as basis for further work on this amendment (attached).

# Workplan

The following list describes items which will be studied/clarified until the 21st WG6 meeting, amongst others:

* Clarify what metadata-only IIF blocks are used for.
* Add window length 768 to the spec text (currently this is only mentioned in the explainer text)
* Clarify whether the signalled delay (and updates to it) is actually applied by the decoder or whether it is meant for the application layer.
* Clarify the decoder interface
  + Why is it possible to signal input\_present = 0 and output\_present = 0, but not input\_present = 1 and output\_present = 1
* Fix typos
  + Syntax of iif\_specific\_config: iif\_predef\_audio\_presentatio\_config --> iif\_predef\_audio\_presentation\_config
* Fix known open technical issues
  + (1==(device\_address\_mask[block\_id]&(1<<d))) --> ((device\_address\_mask[block\_id]&(1<<d)) > 0)
  + Clarify connection between b\_add\_metadata\_block and audio\_signals\_in\_source\_block[idx], i.e. specify that b\_add\_metadata\_block == 1 => audio\_signals\_in\_source\_block[idx] == 0.
* In 4.5.2.17.3.3
  + Clarify how the signature of the current auth sequence is handled, i.e. probably exclude it from the gad\_bytes explicitly without excluding all other signatures not belonging to the current auth sequence.
  + Also clarify which other hashes are included in the calculation of the current signature and which are excluded (especially for the bitrate switching case where there may be more than 1 signature contained in the stream).
* Study further additions for completing and updating the related technologies, with the goal to issue “Text of DAM1” at the 21st WG6 meeting.

# Description of changes

The following gives a high-level overview of use-cases and proposed solutions for AAC-IIF.

## Introduction

While continuously working on enabling various ecoystems for immersive and personalized audio formats (such as MPEG-H Audio or AC-4 Audio) it has become obvious that there is no broadly available solution today for enabling various use-cases related to those experiences in an interoperable way (see more information below). In order to avoid further fragmentation in the industry and to create interoperability, we suggest to create a lightweight MPEG-based solution for those use-cases. It is proposed to introduce an update based on the AAC-LD codec, which can be implemented without major changes to existing core encoders and decoders, and that can serve as an “add-on” technology mostly. With creating an AAC-based interoperable format, we can minimize friction for adoption of the format and therefore eventually enable cost-efficient and seamless updates of existing ecosystems, while also enabling future use-cases in this space.

## Envisioned use-cases

The following Section provides a high-level description of use-cases, which are intended to be addressed by the proposed updates.

### In-Home Distribution

In this use-case, immersive audio needs to be distributed in a home (or home-like) environment (for instance Dolby Atmos content using Dolby Atmos FlexConnect, see [1] and [2]). In the easiest scenario, the rendering is performed on a central device, e.g. a TV-set and then distributed to wireless speakers (either in broadcast mode or via specific speaker feeds). The central device, after a setup procedure, is aware of speaker- and sweet-spot locations. The following figure illustrates this use-case.



1. Centralized Rendering

From the broadcasted signal, the respective speaker must be able to select the parts of the signal which are relevant to it based on simple signaling in an efficient way, and output it accordingly. In order to compensate for changing locations of the speakers (and the sweet-spot) realative to each other, there needs to be a way to dynamically compensate for different delay and gain respectively. In addition, the format needs to be able to handle congested network conditions (packet loss) and needs to implement ways to address those situations.

In another use-case, the speakers are performing the rendering on their own, depending on their location, available drivers, etc. (Decentralized Rendering). The following figure illustrates this use-case.



1. Decentralized Rendering

In order to flexibly handle related metadata for the respective rendering, the format needs to enable transport of metadata related to rendering to support different formats, including relevant MPEG formats (MPEG-H audio and others) but also proprietary formats such as Dolby Atmos.

In those cases, video is usually displayed immediately, while audio needs to be transcoded to be transmitted to speakers (decode on TV, encode for in-home distribution, decode in speaker). Therefore, low audio latency is important in those cases, in order to keep audo and video in sync without having to spend a lot of memory on buffering the video.

In all those cases, the wireless speakers may have listening-capability, (e.g., “smart speakers”), and therefore require echo-management, which requires the speaker to receive one or more echo-references.

### Car Projection

In another use-case, immersive audio is received by a mobile device, for instance via a streaming service, in a car and targeted for playback on the car audio system. To enable a seamless experience, this requires low-latency transmission of audio and video data (e.g. JPEG XS) from the mobile device to the car’s head-unit. The bitrate is typically not very constrained (usually Wifi-based connection). Therefore, a format is needed which enables transmission of high-quality audio at a reasonable bitrate, but with lowest possible latency. Another possibility to realize this use-case is by passing through the bitstream from the phone to the head-unit, and decoding it there. However, this comes with architectural constraints, such as mixing in system sounds and other considerations, and is therefore not optimal in many cases.



1. Car Projection

## Proposed Updates

The immersive Interchange Format (IIF) is an interchange format that enables a reliable, high quality audio transmission via unreliable transport links at low latency. IIF is based on AAC-LD, and the basic characteristics of AAC-LD remain unchanged. AAC-IIF merely provides the possibility to configure it more flexibly, to serve new use-cases.

### Core Codec Updates

As core codec, IIF reuses the coding tools of the MPEG-4 AAC Low delay codec (AAC-LD). In order to enable more flexibility in terms of chosing tradeoffs between latency, computational complexity and quality, we propose to add support for additional frame lengths, i.e., 120/128, 240/256, 384/768 and 960/1024 samples. To accommodate the additional frame lengths, only minor modifications are necessary to the specification and reference software respectively, e.g., definition of scale-factor band and TNS tables.

### Immersive Interchange Format (IIF) Concepts

Immersive Interchange Format (IIF) enables a reliable, high quality audio transmission via unreliable transport links at low latency by constructing the following block-based structures at its core.

#### IIF Block

Table

Description automatically generated with medium confidence

1. IIF Block

A set of audio signals is grouped into a set of audio streams, identified by its stream ID. Audio signal processing typically processes the signals on a frame by frame basis, hence an **IIF Frame** represents a time slice of the entirety of all signals. Within a single frame, a collection of signals belonging to a particular audio stream is defined as an **IIF Block**. The corresponding audio stream is therefore, also referred to as the **IIF Block Stream**, identified by its stream ID. Hence, an IIF Block represents one Frame of an IIF Block Stream and is the basic building block of the IIF bitstream.

#### IIF Block Stream

Table

Description automatically generated

1. IIF Frame having 3 IIF Block Streams - Conceptual

Graphical user interface

Description automatically generated with medium confidence

1. IIF Frame having 3 IIF Block Streams – Bitstream Level

As previously described, an **IIF Block Stream** represents a subset of audio signals within a given set of audio signals, identified by its stream ID. The **IIF Block ID** assigns an IIF Block to its IIF Block Stream, matching its stream ID. IIF Block IDs are strictly contiguous and arranged in ascending order. Since an IIF Frame can consist of signals grouped into several IIF Block Streams, an IIF Frame can therefore, consist of several IIF Blocks. For example, for a given **IIF Frame *N*** having 3 different IIF Block Streams (e.g., stream IDs 0, 1 and 2), there exist 3 different IIF Blocks with the **IIF Block IDs 0, 1 and 2**, respectively.

#### IIF Block Priorities



1. IIF Block Priorities - Conceptual

Table

Description automatically generated with low confidence

1. IIF Block Priorities – Bitstream Level

**IIF Blocks** can represent redundant versions of an IIF **Frame**, either with the identical content or with a lower bitrate. For example, for a given **IIF Frame *N*** and **IIF Block** **ID 0**, two redundant versions of the Block can be created, each being differentiated by their **Priority** **P = 0, 1** and **2**, where 0 denotes the highest priority. In the presence of IIF Blocks having different Priorites, the decoder selects the IIF Block with the highest priority.



1. IIF Block Priorities - Conceptual

Graphical user interface, text, application

Description automatically generated

1. IIF Block Priorities – Bitstream Level

The redundant versions of an IIF Block may be applied to a subset of, or all, IIF Blocks.

#### IIF Access Unit

A grid of yellow squares with red lines

AI-generated content may be incorrect.

1. IIF Access Unit - Conceptual

*A screen shot of a computer

AI-generated content may be incorrect.***Figure 1: IIF Access Unit – Bitstream Level**

An **IIF Access Unit** is a container carrying the *coded representation* of the signals belonging to the current IIF Frame and potentially the redundant versions of previous IIF Frames. Therefore, an IIF Access Unit of Frame N carries the full signal presentation of the current **IIF Frame** **N**, e.g., **IIF Block ID 0** and **1** with **P = 0** and potentially the redundant blocks belonging to previous **IIF Frame** **N-1**, e.g., **IIF Block ID 0, P = 1**.

#### IIF Audio Bus

The **IIF Audio Bus** is a concept that allows mapping of incoming audio signals to either a ***predefined audio presentation***, i.e., channels, objects or HOA, and/or a ***device specific routing*** to be used by devices in selected roles, e.g., as a driver feed/Main or echo reference/Auxiliary. The signals routed to the **IIF Audio Bus** are ordered by their IIF Block ID and the channel order of a given channel layout.

Diagram, schematic

Description automatically generated

1. IIF Audio Bus combined use cases

##### Centralized Rendering Use Case

The IIF Audio Bus only contains a single section for Device Specific Routing.

Diagram, schematic

Description automatically generated

1. IIF Audio Bus Device Specific Routing

###### Device Addressing

For the ***device specific routing***, each IIF Block can be configured to only address a subset of devices by means of a device specific routing matrix. Devices not addressed, can skip over such Blocks.

A picture containing schematic

Description automatically generated

1. Device Adressing

##### Decentralized Rendering Use Case

The IIF Audio Bus only contains a single section for a Predefined Audio Presentation.

Diagram, schematic

Description automatically generated

1. Predefined Audio Presentation

###### Predefined Audio Presentation

Predefined Audio Presentations can be either channels, objects, or HOA

Predefined Audio Presentation – Channels

Channels can signal a channel mode superset that is kept constant for the lifetime of a connection and channel mode subsets which reflect the actual usage of that connection.

Predefined Audio Presentation – Objects

Objects are audio signals associated with metadata. The metadata format is not intendended to be standardized, rather different metadata formats depending on the application of the format can be used.

# Conclusion

It is expected to include the changes described in this document to answer industry demand and to create a standardized interoperable format supporting the described use-cases in a flexible and frictionless way, with a goal of issuing “Text of DAM1” at the 21st WG6 meeting in October 2025.

In addition, we intend to consider further additions for completing and updating the related technologies, with the goal to issue “Text of DAM1” at the 21st WG6 meeting.

# Informative references

1. Dolby Atmos FlexConnect, <https://www.dolby.com/technologies/dolby-atmos/dolby-atmos-flexconnect/>
2. <https://www.tcl.com/us/en/press-releases/tcl-brings-dolby-atmos-flexconnect-2025-qd-mini-led-tvs>
3. Dolby Atmos for Cars, <https://professional.dolby.com/music/dolby-atmos-for-cars/#gref>