Technical Report

ISO/IEC TR 14496-24

Second edition
2025-08

Information technology — Coding of audio-visual objects —

Part 24: Audio and systems interaction

Technologies de l'information — Codage d'objets audiovisuels — Partie 24: Codage audio et interaction de systèmes

Reference number
ISO/IEC TR 14496-24:2025(en)

© ISO/IEC 2025



COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2025

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword ... **iv**
1 Scope ... **1**
2 Normative references ... **1**
3 Terms and definitions ... **2**
4 Motivating audio composition time stamp handling ... **2**
5 AAC Encoder/Decoder Behavior ... **3**
5.1 Example 1: AAC ... 3
5.1.1 Overview ... 3
5.1.2 Pre-roll ... 5
5.1.3 Edit-list ... 5
5.1.4 Compressed Information and decoder behaviour ... 5
5.2 Example 2: HE-AAC ... 6
5.2.1 Overview ... 6
5.2.2 Pre-roll ... 8
5.2.3 Edit-list ... 8
6 Streaming Considerations ... **9**
Annex A (informative) **Relevant ISO Base Media File Format Syntax** ... **10**

© ISO/IEC 2025 – All rights reserved

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives or www.iec.ch/members_experts/refdocs).

ISO and IEC draw attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO and IEC take no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO and IEC had not received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents and https://patents.iec.ch. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see www.iso.org/iso/foreword.html. In the IEC, see www.iec.ch/understanding-standards.

This document was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

This second edition cancels and replaces the first edition (ISO/IEC TR 14496-24:2008), which has been technically revised.

The main changes are as follows:

— addition of details about complex audio and system interaction scenarios and HE-AAC content signalling;

— refactored description of timestamp and delay handling;

— extension of the HE-AAC example.

A list of all parts in the ISO/IEC 14496 series can be found on the ISO and IEC websites.

Any feedback or questions on this document should be directed to the user’s national standards body. A complete listing of these bodies can be found at www.iso.org/members.html and www.iec.ch/national-committees.

© ISO/IEC 2025 – All rights reserved

Information technology — Coding of audio-visual objects —

Part 24:
Audio and systems interaction

1 Scope

This document describes the desired joint behaviour of MPEG-4 Systems (MPEG-4 File Format) and MPEG-4 Audio codecs. It is desired that MPEG-4 Audio encoders and decoders permit finite length signals to be encoded to a file (particularly MPEG-4 files) and decoded again to obtain the identical signal, subject to codec distortions. This enables the use of audio in systems implementations (particularly MPEG-4 Systems), perhaps with other media such as video, in a deterministic fashion. Most importantly, the decoded signal has nothing “extra” at the beginning or “missing” at the end.

This permits:

a) an exact "round trip" from raw audio to encoded file back to raw audio (excepting encoding artefacts);

b) predictable synchronization between audio and other media such as video;

c) correct behaviour when performing random access as well as when starting at the beginning of a stream;

d) identical behaviour when edits are applied in the raw domain and the encoded domain (excepting encoding artefacts).

It is also expected that there be predictable interoperability between encoders (as represented by files) and decoders. There are two kinds of audio "offsets" (or "delay" in the context of transmission): those that are result from the encoding process, and those that are result from the decoding process. This document is primarily concerned with the latter.

These issues are resolved by the following:

— The handling of composition time stamps for audio composition units is specified. Special care is taken in the case of compressed data, like HE-AAC coded audio, that can be decoded in a backward compatible fashion as well as in an enhanced fashion.

— Examples are given that show how a finite length signals can be encoded to an MPEG-4 file and decoded again to obtain the identical signal, excepting codec distortions. Most importantly, the decoded signal has nothing “extra” at the beginning or “missing” at the end.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 14496-3:2019, *Information technology — Coding of audio-visual objects — Part 3: Audio*

ISO/IEC 14496-12, *Information technology — Coding of audio-visual objects — Part 12: ISO base media file format*

© ISO/IEC 2025 – All rights reserved

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 14496-3 and ISO/IEC 14496-12 apply.

ISO and IEC maintain terminology databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at https://www.electropedia.org/

4 Motivating audio composition time stamp handling

Following ISO/IEC 14496-3:2019, subclause 1.6.5.2, there are 3 ways to signal HE-AAC content:

— Implicit signalling (backward compatible): the audioObjectType in the AudioSpecificConfig structure is set to 2, the AudioSpecificInfo does not contain any extension;

— Explicit signalling;

  — Backward compatible: the audioObjectType in the AudioSpecificConfig structure is also set to 2, but the AudioSpecificInfo contains SBR extension;

  — Hierarchical (non-backward compatible): the audioObjectType in the AudioSpecificConfig structure is set to 5.

For compressed data, like HE-AAC coded audio, which can be decoded by different decoder configurations, decoding can be done in a backward-compatible fashion (AAC only) as well as in an enhanced fashion (AAC+SBR). In order to ensure that timestamps are correct (so that audio remains synchronized with other media), the following is taken into consideration concerning MPEG-4 Systems and Audio:

— If compressed data permits both backward-compatible and enhanced decoding, and if the decoder is operating in a backwards-compatible fashion, then the decoder is not expected to take any action. However, if the decoder is operating in enhanced fashion such that it is using a post-processor that inserts some additional delay (e.g. the SBR post-processor in HE-AAC), then it is expected to notify Systems about the additional time delay incurred relative to the backwards-compatible mode. The exact notification mechanism between the audio decoder (post-processor) and the systems layer is implementation-specific and is not reflected in Systems syntax, i.e. the additional delay is not included in the edit list. With the delay thus notified, Systems can compensate for the additional delay.

— Specifically for HE-AAC (using any of the available signalling mechanisms, i.e. implicit signalling, backward compatible explicit signalling, or hierarchical explicit signalling) the original access unit timestamps (i.e. its decoding time stamp, which in this case is the same as its composition time stamp; but also its presentation time, i.e. after application of any edit list instruction) apply to backward-compatible AAC decoding and playback adjustment for delay compensation is expected in case of AAC+SBR decoding.

Figure 1 shows the composition unit that is generated by an AAC decoder (upper half) and by an HE-AAC decoder operating SBR in dual-rate mode (lower half) when being fed with an access unit of an HE-AAC bitstream. Note that the composition time stamp associated to said access unit applies to the n-th sample of the composition unit. For the AAC decoder case, n has the value 1. For the HE-AAC decoder case, n has the value 962+1 to reflect the additional algorithmic delay of 962 samples of the SBR tool at the HE-AAC output sampling rate (which is twice the sampling rate of the backward compatible AAC output).

Note also that in this document, the term "sample" refers to an audio sample and not an ISOBMFF sample, for which the term "access unit" is preferred in this document.

© ISO/IEC 2025 – All rights reserved

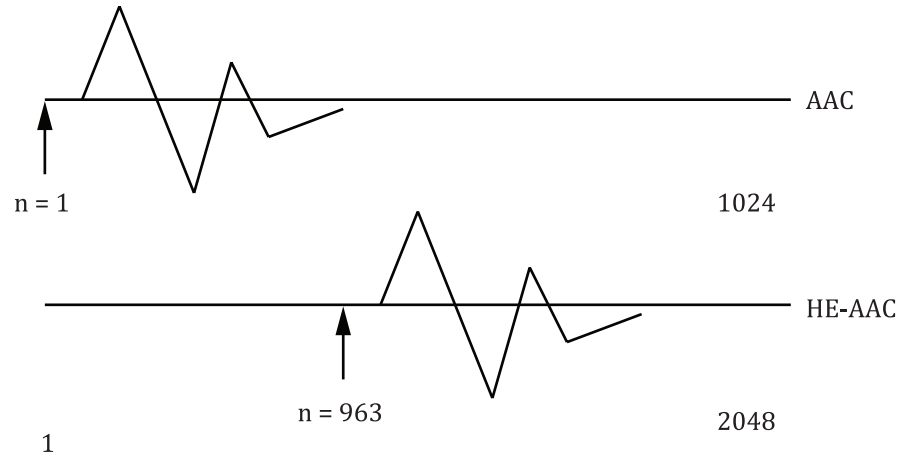


Figure 1 — Composition unit (audio waveform segment) generated by AAC decoder and HE-AAC decoder fed with the same access unit (bitstream frame)

5 AAC Encoder/Decoder Behavior

5.1 Example 1: AAC

5.1.1 Overview

Figure 2 shows the AAC encoder and decoder behaviour with respect to the association of encoder input blocks, access units (AU), timestamps and decoder output blocks or composition units (CU). Note that the input signal is only two and a fraction blocks long (as indicated by the oscillating waveform). The encoder essentially extends the waveform at both ends to facilitate encoding of the entire waveform. The ISO Base Media File Format "helper" information "pre-roll" and "edit-list" facilitate exact reconstruction of the encoded waveform segment in the case that the compressed data is stored in an MPEG-4 Format file.

The specifics of encoder behaviour are:

— The encoder is expected to produce normative access units;

— The timestamp associated with those access units is expected to be the time of the first sample of the waveform in the corresponding composition unit.

© ISO/IEC 2025 – All rights reserved