INTERNATIONAL STANDARD

ISO/IEC 14496-26

First edition
2010-05-01

Information technology — Coding of audio-visual objects —

Part 26: Audio conformance

Technologies de l'information — Codage des objets audiovisuels —

Partie 26: Conformité audio

Reference number
ISO/IEC 14496-26:2010(E)

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

Page

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. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of the joint technical committee is to prepare International Standards. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

ISO/IEC 14496-26 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

This part of ISO/IEC 14496 cancels and replaces:

— ISO/IEC 14496-4:2004, Clause 6,

— ISO/IEC 14496-4:2004/Cor.5,

— ISO/IEC 14496-4:2004/Cor.6,

— ISO/IEC 14496-4:2004/Amd.8:2005, including ISO/IEC 14496:2004/Amd.8:2005/Cor.1:2008,

— ISO/IEC 14496-4:2004/Amd.11:2006, including ISO/IEC 14496-4:2004/Amd.11:2006/Cor.1:2008,

— ISO/IEC 14496-4:2004/Amd.11:2006/Cor.2:2007,

— ISO/IEC 14496-4:2004/Amd.11:2006/Cor.3:2008,

— ISO/IEC 14496:2004-4/Amd.13:2007, including ISO/IEC 14496-4:2004/Amd.13:2007/Cor.1:2007,

— ISO/IEC 14496:2004-4/Amd.13:2007/Cor.2:2007,

— ISO/IEC 14496-4:2004/Amd.14:2007,

— ISO/IEC 14496-4:2004/Amd.15:2007,

— ISO/IEC 14496-4:2004/Amd.18:2007,

— ISO/IEC 14496-4:2004/Amd.19:2007, including ISO/IEC 14496-4:2004/Amd.19:2007/Cor.1:2008,

— ISO/IEC 14496-4:2004/Amd.20:2008, and

— ISO/IEC 14496-4:2004/Amd.22:2008.

ISO/IEC 14496 consists of the following parts, under the general title *Information technology — Coding of audio-visual objects:*

— *Part 1: Systems*

— *Part 2: Visual*

— *Part 3: Audio*

— *Part 4: Conformance testing*

— *Part 5: Reference software*

— *Part 6: Delivery Multimedia Integration Framework (DMIF)*

— *Part 7: Optimised reference software for coding of audio-visual objects*

— *Part 8: Carriage of ISO/IEC 14496 contents over IP networks*

— *Part 9: Reference hardware description*

— *Part 10: Advanced Video Coding*

— *Part 11: Scene description and application engine*

— *Part 12: ISO base media file format*

— *Part 13: Intellectual Property Management and Protection (IPMP) extensions*

— *Part 14: MP4 file format*

— *Part 15: Advanced Video Coding (AVC) file format*

— *Part 16: Animation Framework eXtension (AFX)*

— *Part 17: Streaming text format*

— *Part 18: Font compression and streaming*

— *Part 19: Synthesized texture stream*

— *Part 20: Lightweight Application Scene Representation (LASeR) and Simple Aggregation Format (SAF)*

— *Part 21: MPEG-J Graphics Framework eXtensions (GFX)*

— *Part 22: Open Font Format*

— *Part 23: Symbolic Music Representation*

— *Part 24: Audio and systems interaction* [Technical Report]

— *Part 25: 3D Graphics Compression Model*

— *Part 26: Audio conformance*

— *Part 27: 3D Graphics conformance*

Introduction

ISO/IEC 14496-3 specifies coded representations of audio information. ISO/IEC 14496-3 allows for large flexibility, achieving suitability of ISO/IEC 14496 for many different applications. The flexibility is obtained by including parameters in the bitstream that define the characteristics of coded bitstreams. Examples are the audio sampling frequency bitrate parameters, synchronisation timestamps, the association of bitstreams and synthetic objects within objects.

This part of ISO/IEC 14496 specifies how tests can be designed to verify whether bitstreams and decoders meet the requirements as specified in ISO/IEC 14496-3 and allow interoperability with remote terminals in interactive, broadcast and local (with stored contents) sessions. These tests can be used for various purposes such as

— manufacturers of encoders, and their customers, can use the tests to verify whether the encoder produces bitstreams compliant with ISO/IEC 14496-3,

— manufacturers of decoders and their customers can use the tests to verify whether the decoder meets the requirements specified in ISO/IEC 14496-3 for the claimed decoder capabilities,

— manufacturers and customers of terminals supporting interactive, broadcast and local sessions over a multitude of transport protocols and networks, can use the tests to verify whether the claimed functionalities are compliant with ISO/IEC 14496-6,

— manufacturers of test equipments, and their customers can use the tests to verify compliance with ISO/IEC 14496-3.

Information technology — Coding of audio-visual objects —

Part 26:
Audio conformance

1 Scope

This part of ISO/IEC 14496 specifies how tests can be designed to verify whether compressed data and decoders meet requirements specified by ISO/IEC 14496-3. In this part of ISO/IEC 14496, encoders are not addressed specifically. An encoder may be said to be an ISO/IEC 14496 encoder if it generates compressed data compliant with the syntactic and semantic bitstream payload requirements specified in ISO/IEC 14496-3.

Characteristics of compressed data and decoders are defined for ISO/IEC 14496-3. The compressed data characteristics define the subset of the standard that is exploited in the compressed data. Examples are the applied values or range of the sampling rate and bitrate parameters. Decoder characteristics define the properties and capabilities of the applied decoding process. An example of a property is the applied arithmetic accuracy. The capabilities of a decoder specify which compressed data the decoder can decode and reconstruct, by defining the subset of the standard that may be exploited in the decodable compressed data. Compressed data can be decoded by a decoder if the characteristics of the compressed data are within the subset of the standard specified by the decoder capabilities.

Procedures are described for testing conformance of compressed data and decoders to the requirements defined in ISO/IEC 14496-3. Given the set of characteristics claimed, the requirements that must be met are fully determined by ISO/IEC 14496-3. This part of ISO/IEC 14496 summarises the requirements, cross references them to characteristics, and defines how conformance with them can be tested. Guidelines are given on constructing tests to verify decoder conformance. Some examples of compressed data implemented according to these guidelines are provided as an electronic annex to this document usually together with their uncompressed counterparts (reference waveforms).

2 Normative references

The following referenced documents are indispensable for the application 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 11172-3, *Information technology — Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s — Part 3: Audio*

ISO/IEC 11172-4, *Information technology — Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s — Part 4: Compliance testing*

ISO/IEC 13818-3, *Information technology — Generic coding of moving pictures and associated audio information — Part 3: Audio*

ISO/IEC 13818-4, *Information technology — Generic coding of moving pictures and associated audio information — Part 4: Conformance testing*

ISO/IEC 13818-7, *Information technology — Generic coding of moving pictures and associated audio information — Part 7: Advanced Audio Coding (AAC)*

ISO/IEC 14496-1, *Information technology — Coding of audio-visual objects — Part 1: Systems*

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

ISO/IEC 14496-11, *Information technology — Coding of audio-visual objects — Part 11: Scene description and application engine*

3 Terms and definitions

For the purposes of this document the terms, defintions, symbols and abbreviated terms given in ISO/IEC 14496-1, ISO/IEC 14496-3 and the following apply.

3.1
conformance data
conformance test sequences and **conformance tools**

3.2
conformance tool
tool to check certain conformance criteria

NOTE Conformance tools are provided in the electronic attachments to this part of ISO/IEC 14496.

3.3
conformance test sequence
superset of **compressed data** and its **reference waveforms**

NOTE Examples of conformance test sequences are provided in the electronic attachments to this part of ISO/IEC 14496.

3.4
compressed data
data encoded in accordance with ISO/IEC 14496-3

3.5
reference waveform
decoded counterparts of the **compressed data**

4 Conformance Points

All audio decoders except the LATM-based decoders are part of the MPEG-4 framework. Table 1 gives an overview about the interfaces that have to be provided to test the audio decoders using the MPEG-4 System.

Table 1 — Conformance points

conformance point/interface	data flow direction	description/reference
AudioSpecificConfig	in	audio related decoder specific information, see ISO/IEC 14496-3:2009, (1.6.2.1 AudioSpecificConfig)
audio access units	in	audio related bitstream payload, see ISO/IEC 14496-1:2004 (7.1.2.3 Access Units (AU))
BIFS/AudioSource node	in	see ISO/IEC 14496-11: 2005 (7.2.2.15 Audio Source)
private test info	in	to control some elements which are usually generated by random number generators
audio composition units	out	see ISO/IEC 14496-1: 2004 (7.2.8 Composition Units (CU))

Figure 1 gives an overview about the test bench (MPEG-4 System), the system under test (Audio decoder), and the interfaces between them. Figure 2 gives a more detailed view on the audio decoder, consisting of error protection (EP) decoder and audio core decoder.

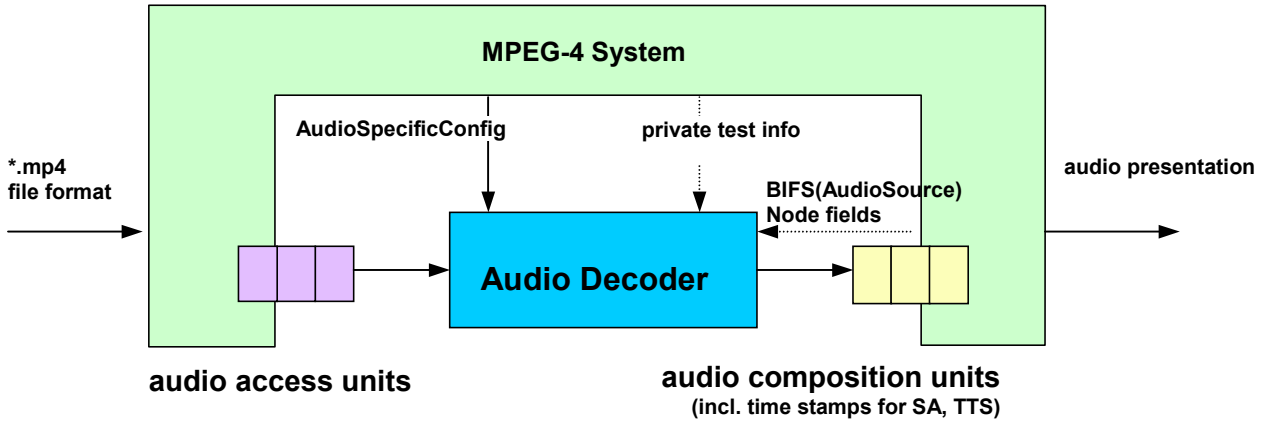

Figure 1 — Audio Conformance Points

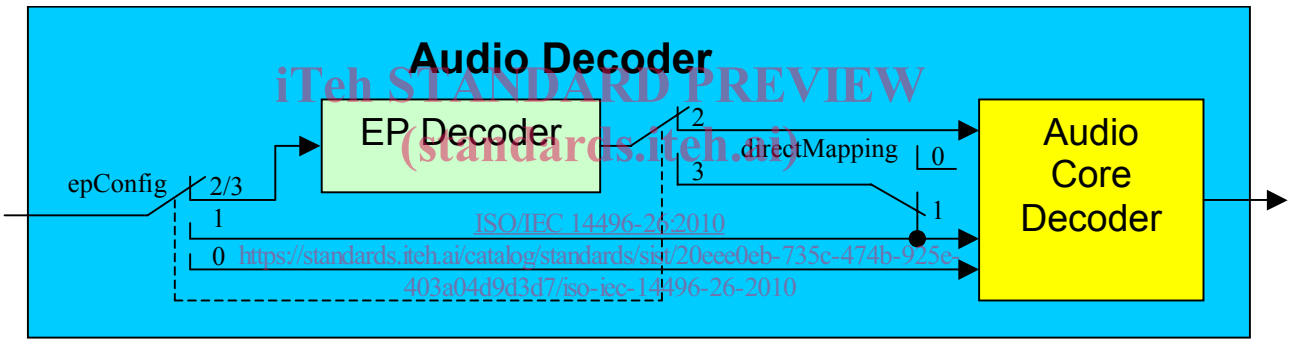

Figure 2 — Audio decoder structure

Clause 7 describes:

The conformance criteria of the audio core decoder.

The conformance criteria of the compressed data not requiring the EP decoder (epConfig == 0 || epConfig == 1).

The properties of the examples of compressed data with (epConfig == 0 || epConfig == 1).

Clause 8 describes:

The conformance criteria of the EP decoder

The conformance criteria of the compressed data requiring the EP decoder (epConfig == 2 || epConfig == 3).

The properties of the examples of compressed data with (epConfig == 2 || epConfig == 3).

Compressed data with different epConfig settings might be available referring to the same reference waveforms. Here, the output of a conforming decoder shall be equal, independently of the used epConfig setting.

For some of the compressed data containing scalable configurations, conformance points are defined at the PCM output of the decoder for *m* layers being decoded from an *n*-layer input, where *m* is an integer in the range 0 (base layer conformance) to *n-1*. The reference PCM decoder output signals corresponding to these conformance points are listed in the respective conformance tables.

5 Profiles

ISO/IEC 14496-3 defines several profiles and several levels within each profile. Conformance is always tested against a certain level within a certain profile. Audio profiles always comprise a set of audio object types. Nevertheless the conformance criteria as described within this document are based on audio object types. The assignment of object types to profiles as well as the level definitions can be found in ISO/IEC 14496-3. The conformance of a certain level within a certain profile is fulfilled, if the conformance of each object type belonging to this profile is fulfilled. The assignment of the provided test sequences to profiles and levels can be found in Clause 12.

6 Conformance data

6.1 File name conventions

For all conformance test sequences, the file name convention given in Table 2 is used.

Table 2 — File name conventions

object type name/ tool name	File Name (compressed)	File Name (uncompressed)
AdvancedAudioBIFS - perceptual apporach	aabper<coreSetup>	-- not applicable --
AdvancedAudioBIFS - physical approach	aabphy<coreSetup>	-- not applicable --
AudioBIFS	ab<coreSetup>_<coder>	ab<coreSetup>_<coder>
AudioBIFS v3	ABv3_<nodeAbbrev><coreSetup>	-- not applicable --
AAC scalable	ac<coreSetup>	ac<coreSetup>[_lay<highestLay>]
AAC LC	al<coreSetup>_<fs>	al<coreSetup>_<fs>[_cut<fac>_boost<facr>] [_level<lvl>][_<chan>]
AAC main	am<coreSetup>_<fs>	am<coreSetup>_<fs>[_cut<fac>_boost<facr>] [_level<lvl>][_<chan>]
AAC LTP	ap<coreSetup>_<fs>	ap<coreSetup>_<fs>
AAC SSR	as<coreSetup>_<fs>	as<coreSetup>_<fs>[_<chan>]
CELP	ce<coreSetup>	ce<coreSetup>[_lay<highestLay>]
ER AAC scalable	er_ac<coreSetup>_ep<epConfig> [<epSetup>]	er_ac<coreSetup>[_lay<highestLay>]
ER AAC LD	er_ad<coreSetup>_<fs>_ep<epConfig> [<epSetup>]	er_ad<coreSetup>_<fs>
ER AAC LC	er_al<coreSetup>_<fs>_ep<epConfig> [<epSetup>]	er_al<coreSetup>_<fs>
ER AAC LTP	er_ap<coreSetup>_<fs>_ep<epConfig> [<epSetup>]	er_ap<coreSetup>_<fs>
SBR (+AAC LC)	al_sbr_<tool>_<fs>_<nchan>[_fsaac<fs>][_sig<sig>]	al_sbr_<mode>_<tool>_<fs>_<nchan> [_fsaac<fs>][_sig<sig>][_<chan>]
SBR (+AAC LC with 960 samples per frame)	al960_sbr_<tool>_<fs>_<nchan> [_fsaac<fs>][_sig<sig>]	al960_sbr_<mode>_<tool>_<fs>_<nchan> [_fsaac<fs>][_sig<sig>][_<chan>]
PS (+SBR+AAC LC)	al_sbr_ps_<coreSetup>	al_sbr_ps_<coreSetup>[_<version>]
SSC	ssc_<tool>_<nchan>[_sig<sig>]	ssc_<mode>_<tool>_<nchan>[_sig<sig>][_<chan>]
DST	dst_<tool>_<nchan>[_sig<sig>]	dst_<mode>_<tool>_<nchan>[_sig<sig>][_<chan>]
Layer-3	l3_<coreSetup>	l3_<coreSetup>

ER BSAC	er_bs<coreSetup>_<fs>_ep<epConfig> [<epSetup>]	er_bs<coreSetup>_<fs>[_lay<highestLay>]
ER CELP	er_ce<coreSetup>_ep<epConfig> [<epSetup>]	er_ce<coreSetup>[_lay<highestLay>]
ER HILN	er_hi<coreSetup>_ep<epConfig> [<epSetup>]	er_hi<coreSetup>[_lay<highestLay>] [_s<speedFac>][_p<pitchFac]
ER HVXC	er_hv<coreSetup>_ep<epConfig> [<epSetup>]	er_hv<coreSetup>[_lay<highestLay>]_<delay>
ER Parametric	er_pa<coreSetup>_ep<epConfig> [<epSetup>]	er_pa<coreSetup>[_lay<highestLay>]_<delay>
ER Twin VQ	er_tv<coreSetup>_ep<epConfig> [<epSetup>]	er_tv<coreSetup>[_lay<highestLay>]
HVXC	hv<coreSetup>	hv<coreSetup>[_lay<highestLay>]_ref<decCfg>
Algorithmic Synthesis and Audio FX	sy<coreSetup>	sy<coreSetup>
TTSI	tts<coreSetup>	tts<coreSetup>
TwinVQ	tv<coreSetup>	tv<coreSetup>[_lay<highestLay>]
ALS	als_<tool>_<coreSetup>	als_<tool>_<coreSetup>
SLS	sls<coreSetup>_<fs>_<bitres>	sls<coreSetup>_<fs>_<bitres>
Layer-1	l1_<coreSetup>	l1_<coreSetup>
Layer-2	l2_<coreSetup>	l2_<coreSetup>
ER AAC ELD	er_eld<coreSetup>_<fs>_ep<epConfig> [<epSetup>]	er_eld<coreSetup>_<fs>

<bitres> can be 16 or 24 and indicates the bit resolution of the coded wavefile

<chan> indicates the channel for multi-channel sequences (f<number> - number of the front channel, b<number>- number of the back channel, s<number> - number of the side channel, l<number> - number of the LSF channel).

<coder> indicates the coder used to encode the content (ce – CELP, sa – Structured Audio, pcm – PCM)

<coreSetup> refers to a certain audio coder setup. It is most likely a number, but might also contain characters.

<delay> refers to the decoder delay, it can become “ld” (low delay) or “nd” (normal delay).

<epConfig> can be 0, 1, 2 or 3, depending on epConfig (defined in AudioSpecificConfig).

<epSetup> is required if (epConfig==2 || epConfig==3). It refers to a certain error protection setup.

<fs> sampling frequency (08, 11, 12, 16, 22, 24, 32, 44, 48, 64, 88 or 96).

_level<lvl> refers to the level with regard to DRC.

_cut<fac>_boost<fac> referes to the cut and boost factors with regard to DRC.

_lay<highestLay> is required for any scalable configuration. It marks the highest layer of the scalable configuration used for decoding (starting with 0 for the core layer).

_p<pitchfac> is a number refering to the decoder configuration with regard to the pitch factor.

_ref<decCfg> is a number refering to the decoder configuration with regard to delay mode, speed and pitch change.

_s<speedfac> is a number refering to the decoder configuration with regard to the speed factor.

<tool> indicates the SBR module mainly targeted by the test sequence. Possible values are “e” for testing the envelope adjuster “s” for testing sine addition, “gh” for testing time-grid transitions in combination with changes of SBR header data, “i” for testing inverse filtering, “qmf” for testing the QMF implementation, “cm” for testing various channel modes, “sig” for testing SBR signaling, “twi” for QMF identification, and “sr” for testing various combinations of sampling rates.

<nodeAbbrev> is the abbreviation of one of the AudioBIFS v3 node names.

<nchan> corresponds to the number of channels present in the conformance test sequence. It is either a single integer, in which case it refers to the number of main audio channels, or two integers separated by a “.”, in which case the first integer equals the number of main audio channels, while the second number equals the number of low frequency enhancement channels.

fsaac<fs> corresponds to the sampling rate of the underlying AAC-LC data. If it is omitted, it is half the sampling rate given as output sampling rate.

<sig> is an integer describing the kind of signalling used according to the table below. If this value is omitted, backwards compatible explicit signalling of SBR is used.

file name conventions

sig	Signalling method used
0	Implicit signalling of SBR
1	Hierarchical explicit signalling of SBR
2	Backwards compatible explicit signalling of SBR

<mode> is either “hq” or “lp” for the high quality or the low power version of the SBR decoding algorithm respectively.

<version> is either “bl” or “ur” for the baseline or the unrestricted version of the parametric stereo decoding algorithm respectively.

With respect to file extensions, the following rules are applied:

Compressed	MPEG-4 file format	.mp4
Compressed	native MPEG-1/2 Audio storage format	.mpg
Compressed	Audio data interchange format	.adif
Compressed	Audio data transport stream	.adts
Compressed	AudioSyncStream	.ass
Compressed	EPAudioSyncStream	.ess
Compressed	AudioPointerStream	.aps
Uncompressed	HILN Conformance Test Parameters	.ctp
Uncompressed	WAVE format (uncompressed PCM format)	.wav
Uncompressed	TTSI decoded text and control digits	.txt

6.2 Content

The test set includes a set of sine sweeps, a set of musical/speech test sequences and a set of noise-like test sequences. The supplied sine sweeps with an amplitude of -20dB relative to full scale have an absolute amplitude of +/- 0.1.

7 Audio Object Types

7.1 General

This Clause lists all audio object types. It starts with a general description, which may be related to more than one object type.

This Clause contains general descriptions for conformance testing on compressed data and decoders. Unless explicitly restricted, these descriptions are related to all object types.

7.1.1 Compressed Data

7.1.1.1 Characteristics

Characteristics of compressed data specify the constraints that are applied by the encoder in generating the compressed data. These syntactic and semantic constraints may, for example, restrict the range or the values of parameters that are encoded directly or indirectly in the compressed data. The constraints applied to a given compressed data may or may not be known a priori.

Decoder relevant compressed data may consist of the following parts:

decoder specific information (AudioSpecificConfig)

BIFS/AudioSource node (field information)

audio access units (establishing the bitstream payload)

7.1.1.1.1 ESC instance configuration

In case of epConfig=1, each instance of each sensitivity category belonging to one frame is stored separately within a single access unit, i.e. there exist as many elementary streams as instances defined within a frame.

Note: In case of epConfig=3, the mapping between EP classes and ESC instances is signaled by the data element directMapping. In case of directMapping=1, the restrictions regarding the ESC instance configuration apply accordingly to the EP class configuration.

The following table gives an overview about the valid configurations: