INTERNATIONAL STANDARD

ISO/IEC 13818-4

First edition
1998-12-01

AMENDMENT 1
1999-06-15

Corrected and reprinted
2000-04-01

Information technology — Generic coding of moving pictures and associated audio information —

Part 4:
Conformance testing

AMENDMENT 1: Advanced Audio Coding (AAC) conformance testing

Technologies de l'information — Codage générique des images animées et des informations sonores associées —

Partie 4: Essais de conformité

AMENDEMENT 1: Essai de conformité de codage audio avancé

Reference number
ISO/IEC 13818-4:1998/Amd.1:1999(E)

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. 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.

Amendment 1 to ISO/IEC 13818-4:1998 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

Information technology — Generic coding of moving pictures and associated audio information — Part 4: Conformance testing

AMENDMENT 1

Advanced Audio Coding (AAC) conformance testing

1) Add the following reference to subclause 1.2 Normative References:

ITU-R Document TG10-2/3- E only, " Basic Audio Quality Requirements for Digital Audio Bit-Rate Reduction Systems for Broadcast Emission and Primary Distribution", 28 October 1991.

2) Replace the following definitions in subclause 2.1 with the following:

2.1.142 low frequency enhancement (LFE) channel: A limited bandwidth channel for low frequency audio effects in a multichannel system.

2.1.151 modified discrete cosine transform (MDCT): A transform which has the property of time domain aliasing cancellation. An analytical expression for the MDCT can be found in B 2.3.1.2.

2.1.258 variable length code (VLC): A code word assigned by variable length encoder (see variable length coding).

2.1.259 variable length decoder (VLD): A procedure to obtain the symbols encoded with a variable length coding technique.

3) Add the following definitions in subclause 2.1 as indicated:

2.1.145.1 main audio channels: All single_channel_elements (see 3.2.1) or channel_pair_elements (see 3.2.1) in one program.

2.1.193.1 program [AAC]: A set of main audio channels, coupling_channel_elements (see 3.2.1), lfe_channel_elements (see 3.2.1), and associated data streams intended to be decoded and played back simultaneously. A program may be defined by default (see 3.5.1) or specifically by a program_configuration_element (see 3.2.1). A given single_channel_element (see 3.2.1), channel_pair_element (see 3.2.1), coupling_channel_element, lfe_channel_element or data channel may accompany one or more programs in any given bitstream.

2.1.214.1 Sampling Frequency (Fs): Defines the rate in Hertz which is used to digitize an audio signal during the sampling process.

2.1.230.1 spectral coefficients: Discrete frequency domain data output from the analysis filterbank.

2.1.259.1 variable length encoder: A procedure to assign variable length codewords to symbols.

4) Add the following Arithmetic Operator in subclause 2.2.1 following the operator

Sign(x) = -1 when x<0:

INT ( ) Truncation to integer operator. Returns the integer part of the real-valued argument.

5) Add the following as subclause 2.6 for conformance testing of ISO/IEC 13818-7:1997, Information technology — Generic coding of moving pictures and associated audio information — Part 7: Advanced Audio Coding (AAC):

2.6 Advanced Audio Coding (AAC)

2.6.1 Introduction

ISO/IEC 13818-7 describes the Advanced Audio Coding (AAC) system for multichannel audio. This part of ISO/IEC 13818 includes much flexibility to enable its use in many different applications. The flexibility is obtained by specifying a toolset that contains several required and optional tools, as well as several parameterized elements in the bitstream.

This subclause of the specification specifies how tests can be designed to verify whether bitstreams and decoders meet the requirements as specified in ISO/IEC 13818-7. 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 valid bitstreams for a particular AAC profile.
- manufacturers of decoders and their customers can use the tests to verify whether the decoder meets the requirements specified in ISO/IEC 13818-4 for the claimed decoder capabilities.

2.6.2 Scope

This subclause of this part of ISO/IEC 13818 specifies how tests can be designed to verify whether bitstreams and decoders meet requirements specified in ISO/IEC 13818-7. In this part of ISO/IEC 13818, encoders are not addressed specifically. An encoder may be said to be an ISO/IEC 13818-7 encoder if it generates bitstreams compliant with the syntactic and semantic bitstream requirements specified in ISO/IEC 13818-7.

Characteristics of coded bitstreams and decoders are defined for ISO/IEC 13818-7. The characteristics of a bitstream define the subset of the standard that is exploited in the bitstream. 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 coded bitstreams the decoder can decode and reconstruct, by defining the subset of the standard that may be exploited in decodable bitstreams. A bitstream can be decoded by a decoder if the characteristics of the coded bitstream are within the subset of the standard specified by the decoder capabilities.

Procedures are described for testing conformance of bitstreams and decoders to the requirements defined in ISO/IEC 13818-7. Given the set of characteristics claimed, the requirements that must be met are fully determined by ISO/IEC 13818-7. This subclause summarizes the requirements, cross references them to characteristics, and defines how conformance with them can be tested. Guidelines are given on constructing tests to verify bitstream and decoder conformance. This part of ISO/IEC 13818 gives guidelines on how to construct bitstream test suites to check or verify decoder conformance. In addition, some test bitstreams implemented according to those guidelines are provided as an electronic annex.

2.6.3 AAC Bitstream Characteristics

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

Encoders may apply restrictions to the following parameters of the bitstream:

a) sampling frequency
b) mono_mixdown_element
c) stereo_mixdown_element
d) use of prediction in main profile
e) pulse_data
f) window_shape
g) program_config_element
h) M/S stereo
i) intensity stereo
j) TNS
k) data_stream_element
l) dependently switched coupling channel

m) independently switched coupling channel
n) LFE channel
o) matrix-downmix

2.6.4 Decoder Characteristics

2.6.4.1 Profile Specification

The naming convention for ISO/IEC 13818-7 AAC decoders dictates that a decoder be specified as an A.L.I.D Channel <Profile Name> Profile ISO/IEC 13818-7 AAC Decoder, where A is replaced by the number of main audio channels, L by the number of LFE channels, I by the number of independent coupling channels, D by the number of dependently switched coupling channels, and Profile Name by the actual profile name (Main, Low-Complexity, or Scaleable Sampling Rate). An example would be a 5.1.1.1 Channel Main Profile ISO/IEC 13818-7 AAC Decoder, indicating a decoder capable of decoding 5 main audio channels, one LFE channel, one independent coupling channel, and one dependent coupling channel, with each of the channels using the profile specified. This can be abbreviated as M.5.1.1.1. Similarly, a Low Complexity decoder can be specified by a leading "L", and an SSR profile by an "S".

A conforming decoder must support a minimum capability in terms of the number of main audio channels, LFE channels, independent coupling channels, and dependent coupling channels as specified below:

Table 2-8 Minimum Decoder Capability for 1, 2, 3, 4, 5, and 7 Main Audio Channels vs. Profile

Number of Main Audio Channels	Main Profile Capability	Low Complexity Profile Capability	SSR Profile Capability
1	1.0.0.0	1.0.0.0	1.0.0.0
2	2.0.0.0	2.0.0.0	2.0.0.0
3	3.0.1.0	3.0.0.1	3.0.0.0
4	4.0.1.0	4.0.0.1	4.0.0.0
5	5.1.1.1	5.1.0.1	5.1.0.0
7	7.1.1.2	7.1.0.2	7.1.0.0

Note that a decoder may support more than the number of channel elements specified in the table above and still be a conformant decoder. Thus, M.5.1.1.1, M.5.1.1.2, L.5.1.1.1,and S.5.1.0.0 are all conformant 5 channel configurations. However, M.5.1.0.0, L.5.1.0.0, and S.5.0.0.0 are not conformant 5 channel configurations.

2.6.4.2 Decoder Modifications

A conforming decoder may also support any of the following modifications of some parameters in audio bitstreams:

Bitstream Characteristic	Normative Clause	Variation
sampling rate	3	a decoder may support only a subset of possible sampling rates, decoder manufacturers must specify the rates which are supported
profile	2.1	a decoder may support additional channel elements beyond the minimums listed for its profile
program configuration	3.5	a decoder is only required to decode one program of a multi-program bitstream
data_stream_element	3.6	a decoder is not required to store or present data recovered from data_stream_elements
mono-mixdown element	3.3.8	a decoder is not required to present audio from the mono-mixdown element

stereo-mixdown element	3.3.8	a decoder is not required to present audio from the stereo-mixdown element
matrix-mixdown	3.3.8	a decoder is not required to calculate a matrix-mixdown signal

2.6.4.3 Decoder Buffer Requirements

Refer to 3.2.2 of ISO/IEC 13818-7.

2.6.5 Procedure to Test Bitstream Conformance

Each bitstream shall meet the syntactic and semantic requirements specified in ISO/IEC 13818-7. This subclause describes a set of semantic tests to be performed on bitstreams. The procedure to verify whether the syntax is correct is straightforward and therefore not defined in this subclause. In the description of the semantic tests it is assumed that the tested bitstream contains no errors due to transmission or other causes. For each test the condition or conditions that must be satisfied are given, as well as the prerequisites or conditions in which the test can be applied. Note that the application of these tests requires parsing of the bitstream to the appropriate levels, which in some cases goes as far as the spectral_data recovery.

2.6.5.1 Parsing an ADIF header

adif_id: shall be encoded with the value 0x41444946, the ASCII representation of the string „ADIF“.

2.6.5.2 Parsing ADTS headers

2.6.5.2.1 adts_fixed_header

syncword: shall be encoded with the binary value 1111 1111 1111.

ID: shall not be encoded with the value 0.

layer: shall be encoded with the binary value 00.

profile: shall not be encoded with the binary value 11.

sampling_frequency_index: shall be encoded with a value no greater than 0xb.

2.6.5.2.2 adts_variable_header

frame_length: shall be encoded with the length of the frame, including headers and error check (if present)

2.6.5.3 Decoding of raw data blocks

id_syn_ele: if a program configuration element (PCE) is present, it must be the first syntactic element in a raw_data_block, indicated by id_syn_ele encoded with a value of ID_PCE

element_instance_tag: ensure that element_instance_tag numbers within each element type are unique within each frame. This restriction does not apply to data_stream_elements (DSE), which may have duplicated element_instance_tags.

2.6.5.4 Decoding an individual_channel_stream

ics_reserved_bit: must be set to zero

max_sfb: must be <= num_swb_long or num_swb_short as appropriate for window_sequence and sampling frequency

2.6.5.5 Noiseless Coding

sect_cb[g][i]: shall not be encoded with the binary values 1100 or 1101.

Intensity codebooks INTENSITY_HCB and INTENSITY_HCB2 shall not occur in a single_channel_element, the left channel of a channel pair element, a coupling channel element, or an LFE. Intensity codebooks can only occur in a channel_pair_element if the common_window field is set to 1.

sect_len_incr: the sum of all sect_len_incr elements for a given window group shall equal max_sfb.

hcod_sf[ ]: shall only be encoded with the values listed in the scalefactor Huffman table, Table A.1.

hcod[sect_cb[g][i]][w][x][y][z]: shall only be encoded with the values listed in Huffman codebooks 1, 2, 3, or 4.

hcod[sect_cb[g][i]][y][z]: shall only be encoded with the values listed in Huffman codebooks 5 thru 11.

hcod_esc_y: shall be encoded with a value no larger than 8191, i.e., it shall be encoded with an initial escape sequence consisting of no more than nine '1' bits followed by an escape separator of '0'.

hcod_esc_z: shall be encoded with a value no larger than 8191, i.e., it shall be encoded with an initial escape sequence consisting of no more than nine '1' bits followed by an escape separator of '0'.

pulse_data_present: shall be encoded with a value of 0 when window_sequence is EIGHT_SHORT_SEQUENCE.

pulse_start_sfb: shall be smaller than num_swb_long_window[fs_index].

pulse_offset[i]: swb_offset_long_window[pulse_start_sfb] + pulse_offset[0] + ... + pulse_offset[number_pulse] must be no greater than 1023.

pulse_amp[i]: shall be encoded with a value small enough such that the compensated quantized spectral coefficient is no greater than 8191.

2.6.5.6 Scalefactors

hcod_sf[ ]: shall only be encoded with the values listed in the scalefactor Huffman table, Table A.1.

2.6.5.7 Joint Coding

2.6.5.7.1 MS Stereo

ms_mask_present: shall not be encoded with the binary value 11.

2.6.5.7.2 Intensity Stereo

hcod_sf[ ]: shall only be encoded with the values listed in the scalefactor Huffman table, Table A.1.

Intensity codebooks INTENSITY_HCB and INTENSITY_HCB2 shall not occur in a single_channel_element, the left channel of a channel pair element, a coupling channel element, or an LFE. Intensity codebooks can only occur in a channel_pair_element if the common_window field is set to 1.

2.6.5.8 Coupling Channel

The number of dependently-switched and independently-switched coupling channel elements must not exceed the allowed numbers specified by the profile naming convention (see 0, Decoder Characteristics).

ind_sw_cce_flag: shall not be encoded with the binary value of 1 if independently-switched coupling channel elements are not specified by the profile naming convention (see 0, Decoder Characteristics).

num_coupled_elements: shall not be encoded with a value greater than the total number of single_channel_elements and channel_pair_elements.

cc_target_is_cpe: shall be encoded with the binary value 1 if the syntactic element with element_instance_tag of cc_target_tag_select is a channel_pair_element; otherwise, it shall be encoded with the binary value of 0.

cc_target_tag_select: shall only be encoded with a binary value equal to the element_instance_tag of a single_channel_element or a channel_pair_element of the current frame.

2.6.5.9 Prediction

predictor_data_present: shall not be encoded with the binary value of 1 for low-complexity or SSR profiles.

predictor_reset: shall be encoded with the binary value of 1 for at least one out of every eight consecutive frames for programs in which predictor_data_present is 1.

predictor_reset_group_number: shall not be encoded with the binary values 00000 or 11111; every valid value shall occur once within a consecutive series of 30 occurrences of pred_reset_group_number.

2.6.5.10 Temporal Noise Shaping

length[w][filt]: must be small enough such that the lower bound of the filtered region, indicated by 'bottom', does not exceed the start of the array containing the spectral coefficients (spec[w])

order[w][filt]: must not exceed the maximum permitted order depending on the specified profile (see 2.1.1 of ISO/IEC 13818-7)

2.6.5.11 Filterbank and Block Switching

window_sequence: The meaningful window_sequence transitions are as follows:

from ONLY_LONG_SEQUENCE to	{ONLY_LONG_SEQUENCE LONG_START_SEQUENCE
from LONG_START_SEQUENCE to	{EIGHT_SHORT_SEQUENCE LONG_STOP_SEQUENCE
from LONG_STOP_SEQUENCE to	{ONLY_LONG_SEQUENCE LONG_START_SEQUENCE
from EIGHT_SHORT_SEQUENCE to	{EIGHT_SHORT_SEQUENCE LONG_STOP_SEQUENCE

Other, non-meaningful, window_sequence transitions are also possible:

from ONLY_LONG_SEQUENCE to	{EIGHT_SHORT_SEQUENCE LONG_STOP_SEQUENCE
from LONG_START_SEQUENCE to	{ONLY_LONG_SEQUENCE LONG_START_SEQUENCE
from LONG_STOP_SEQUENCE to	{EIGHT_SHORT_SEQUENCE LONG_STOP_SEQUENCE
from EIGHT_SHORT_SEQUENCE to	{ONLY_LONG_SEQUENCE LONG_START_SEQUENCE

A conformant bitstream must consist of only meaningful window_sequence transitions. However, decoders are required to handle non-meaningful window_sequence transitions as well. Test bitstreams L3 and S17 are provided respectively for Main, Low-Complexity, and Scaleable Sampling Rate profiles to test decoder performance on non-meaningful transitions. The performance requirements for non-meaningful window_sequence transitions are the same as for the meaningful transitions.

2.6.5.12 Gain Control

aloccode: must satisfy the following conditions:

$$\text{aloccode}[B][w][m_1] < \text{aloccode}[B][w][m_2], \quad 1 \le m_1 \le m_2 \le \text{adjust\_num}[B][w]+1$$

where B is the Band ID, an integer between 1 and 3, and w is the Window ID, an integer from 0 to 7.

2.6.5.13 LFE

The number of LFEs must not exceed the allowed number specified by the profile naming convention (see 2.6.4, Decoder Characteristics).

The window_shape field of any LFE shall always be encoded with a value of 0 (sine window).

The window_sequence field of any LFE shall always be encoded with a value of ONLY_LONG_SEQUENCE.

The highest non-zero spectral coefficient of any LFE shall be 12.

The predictor_data_present_flag of any LFE shall be encoded with a value of 0.

Temporal noise shaping shall not be used in any LFE.

2.6.5.14 Program Configuration Elements

Any program may contain no more main audio channels, LFE channels, independent coupling_channel_elements, and dependent coupling_channel_elements than specified by the profile naming convention (see 0, Profile Specification).

mono_mixdown_element_number: must be encoded with the element_instance_tag of a single_channel_element.

stereo_mixdown_element_number: must be encoded with the element_instance_tag of a channel_pair_element.

matrix_mixdown_idx_present: shall only be encoded with a value of 1 if a 3 front/2 rear 5-channel program is indicated for this PCE.

front_element_tag_select: must be encoded with the element_instance_tag of either a single_channel_element or a channel_pair_element.

side_element_tag_select: must be encoded with the element_instance_tag of either a single_channel_element or a channel_pair_element.

back_element_tag_select: must be encoded with the element_instance_tag of either a single_channel_element or a channel_pair_element.

lfe_element_tag_select: must be encoded with the element_instance_tag of a lfe_channel_element.

assoc_data_element_tag_select: must be encoded with the element_instance_tag of a data_stream_element.

cc_element_is_ind_sw: must be encoded with the same value as the ind_sw_cce_flag field of the coupling_channel_element corresponding to valid_cc_element_tag_select.

valid_cce_element_tag_select: must be encoded with the element_instance_tag of a coupling_channel_element.

2.6.6 Procedure to Test Decoder Conformance

To test audio decoders, this part of ISO/IEC 13818 includes a number of test sequences. Supplied sequences cover Main, Low Complexity, and Scaleable Sampling-Rate profile decoders and are provided for sampling rates of 8, 11.025, 12, 16, 22.05, 24, 32, 44.1, 48, 64, 88.2, and 96 kHz. The test set includes a set of sine sweeps, as listed in Table 2-9, and a set of musical test sequences, as listed in Tables 2-10 through 2-12. The extension_fs is appended to the bitstream name to indicate the sampling rate of the test sequence. Possible values of fs are 8, 11, 12, 16, 22, 24, 32, 44, 48, 64, 88, and 96, corresponding to the possibly non-integer sampling rates listed above. For each bitstream, two bitrates are listed in the table. The lower bitrate is to be used for sampling rates of 16kHz and below, and the higher bitrate is to be used at sampling rates above 16kHz.

The following test procedure applies to all sine sweep signals: Testing is done by comparing the output of a decoder under test with a reference output also supplied by this part of ISO/IEC 13818 using the procedure described in section 2.6.6.1. Software is provided for performing this verification procedure. Measurements are carried out relative to full scale where the output signals of the decoders are normalized to be in the range between -1 and +1.To be called an ISO/IEC 13818-7 audio decoder, the decoder shall provide an output such that the rms level of the difference signal between the output of the decoder under test and the supplied reference output is less than 2^{-15}/sqrt(12). In addition, the difference signal shall have a maximum absolute value of at most 2^{-14} relative to full-scale.

This test only verifies the computational accuracy of an implementation.

For the remaining test sequences, a check of conformance using the LSB criterion or other measurements (e.g. objective perceptual measurement systems) is not mandatory, but highly recommended. This also applies to bitstreams with non-meaningful window sequences.

2.6.6.1 Calculation of RMS

All measurements are carried out relative to full scale where the output signals of the decoder and supplied test sequences are normalized to be in the range between -1.0 and +1.0. The supplied sine sweeps, listed in 0, with an amplitude of -20dB relative to full scale have an absolute amplitude of +/-0.1.

The test sequences have a precision (P) of 24 bits, where the most significant bit (MSB) will be labeled bit 0 and the least-significant bit (LSB) will be labeled bit 23. The most significant bit (bit 0) represents the value of -1, the second most significant bit (bit 1) represents the value of +1/2, etc.

$$\text{value of bit 0 (MSB)} = -\frac{1}{2^0} = -1$$

$$\text{value of bit 1} = \frac{1}{2^1} = \frac{1}{2}$$

$$\text{value of bit 2} = \frac{1}{2^2} = \frac{1}{4}$$

$$\vdots$$

$$\text{value of bit 23 (LSB)} = \frac{1}{2^{23}} = \frac{1}{8{,}388{,}608}$$

The output signal of the decoder under test is required to be in the same format. In the case that the output of the decoder has a precision of P' bits and if P' is smaller than 24, then the output is extended to 24 bits by setting bit P' through bit 23 to zero. In the next step, the difference (*diff*) of the samples of these signals has to be calculated. Every channel of a multichannel bitstream shall be tested. The total number of samples for each channel is N.

diff(n) = ' output signal of decoder under test *(n)*' - ' supplied test sequence *(n)*', for $n = 1$ to N

diff(n) = ' output signal of decoder under test *(n)*' - ' supplied test sequence *(n)*', for $n = 1$ to N

The values of all difference samples shall be squared, summed, divided by N and then the square-root shall be calculated. This calculation finally gives the rms level.

$$rms = \sqrt{\frac{1}{N}\sum_{n=1}^{N} diff(n)^2}$$

The decoder under test may be called an ISO/IEC 13818-7 audio decoder if the rms is less than 1/(2^15 * 12^0.5) and if the maximum absolute value of *diff* is less than or equal to 1/2^14.

2.6.9.2 Descriptions of the ISO/IEC 13818-7 (AAC) audio test bitstreams

Table 2-9 Sine Sweeps

	SIN1_fs	SIN2_fs	SIN3_fs
profile	Main	Low-Complexity	SSR
bitrate	40/64	40/64	40/64
# single channel elements	1	1	1
# channel pair elements	0	0	0
# LFE channels	0	0	0
prediction	Yes	No	No

Table 2-10 Main Profile Bitstreams

	M1_fs	M2_fs	M3_fs
bitrate	40/64	80/128	240/512
# single channel elements	1	0	1
# channel pair elements	0	1	2
# LFE channels	0	0	1
# Dep coupling channels	0	0	1
# Indep Coupling Channels	0	0	2
intensity		Yes	Yes
MS		Yes	Yes
prediction	Yes	Yes	Yes
window shape		Yes	Yes
tns		Yes	Yes
pulse data			
Multi program (2)			Yes
Arithmetic torture		Yes	
Nonmeaningful window_sequence transitions			

All Main profile Decoders must also be able to decode all Low Complexity Profiles.